

Original Research Article**DOI - 10.26479/2017.0301.01**

WIDESPREAD OCCURRENCE OF CPK-1 PSEUDOKNOTS AND RELATED RNA STRUCTURES

Jiang Zhu¹, Xiaolan Huang², Zhihua Du^{1*}

1.Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, IL 62901, USA.

2.Department of Computer Science, Southern Illinois University at Carbondale, IL 62901, USA.

ABSTRACT: Many naturally occurring or SELEX RNA pseudoknots may share a common set of unique structural features that define the previously proposed CPK-1 (common pseudoknot motif 1) pseudoknot family. A CPK-1 pseudoknot is formed when a 3'- strand of RNA binds asymmetrically to 6 or 7 nucleotides within the loop region of a hairpin, leaving only 1-2 unpaired nucleotides at the 5'-end of the loop. In the tertiary structure, the two stems (stem1 and 2) of the pseudoknot stack coaxially to form a quasicontinuous helix; the 1-2 unpaired nucleotides in loop1 cross the major groove of stem2, with their bases being embedded inside the major groove. By exquisitely exploiting the characteristic properties of the A-form RNA helix, the CPK-1 fold maximizes base pairing and stacking interactions while minimizes exposure of the hydrophobic bases to solvent. The conserved structural features originally observed in CPK-1 pseudoknots can also mediate other types of RNA interactions, including long-distance intramolecular interactions and intermolecular interactions. Several putative cases of such interactions are presented. RNA molecules harbouring the conserved CPK-1 structural features assume a wide variety of biological functions. We also suggest computational methods for the detection of these kinds of RNA structures. We further propose a specific antisense strategy based on forming intermolecular CPK-1-like structures.

KEYWORDS: RNA structure, CPK-1, pseudoknot ,

***Corresponding Author: Dr. Zhihua Du Ph.D.**

Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, IL 62901,
USA

* Email Address: zdu@chem.siu.edu

1. INTRODUCTION

Wide-spread occurrence of the CPK-1 pseudoknots

Pseudoknots are broadly defined as RNA structures that are formed when a stretch of nucleotides within a loop region basepairs with a complimentary sequence outside that loop (1,2). If the loop is in a hairpin, the pseudoknot is called an H-type pseudoknot (Figure 1). All pseudoknots must contain two helical stems (S1 and S2) and two non-equivalent loops (L1 and L2, crossing the major groove and minor groove of S2 and S1 respectively). A third loop (L3) is optional. In the absence of L3, S1 and S2 can stack coaxially (Figure 1B). Pseudoknots assume various biological functions, the best known of which is to stimulate efficient -1 ribosomal frameshifting or stop codon readthrough in RNA viruses (3) -1 ribosomal frameshifting and readthrough are two of the major translational recoding mechanisms utilized by some RNA viruses to express their structural and enzymatic proteins at a defined ratio (4,5). Efficient recoding usually requires an RNA pseudoknot that locates several nucleotides downstream from the recoding site (the 0 frame stop codon in readthrough or the slippery sequence in -1 ribosomal frameshifting). To assess the strategic importance of the recoding pseudoknots within the viral genomic RNAs, we had carried out a genome-wide analysis of frameshift and readthrough-

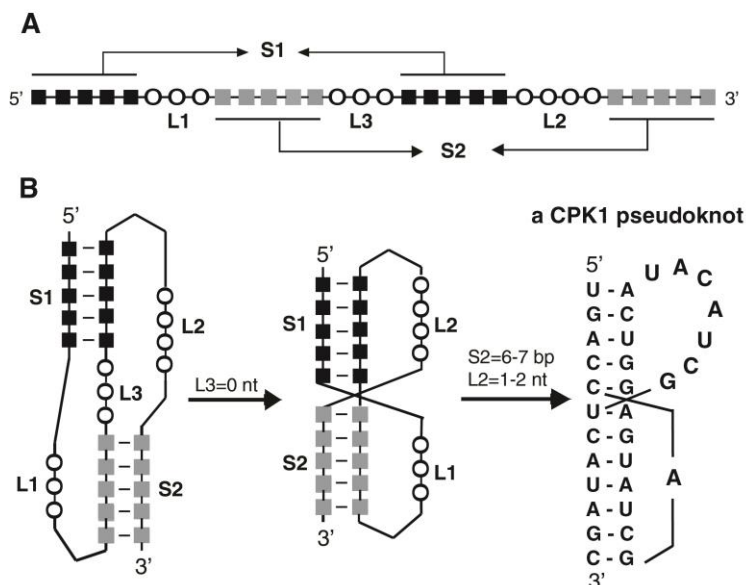


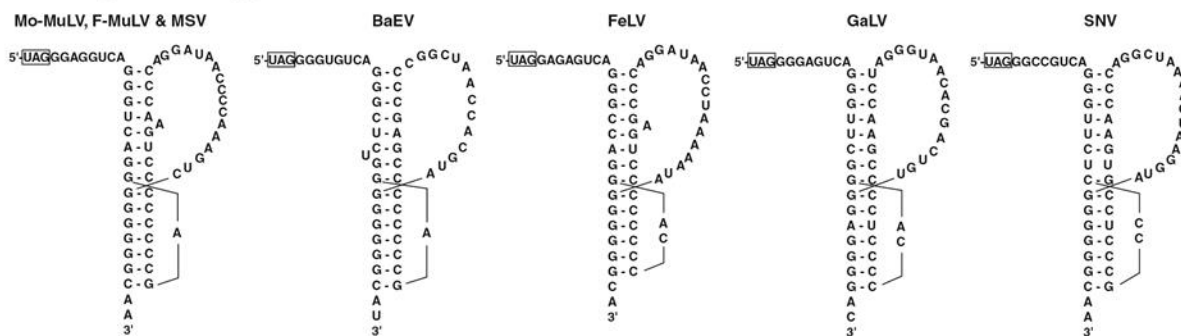
Figure 1: The H-type RNA pseudoknots. Abbreviations: S1&2, stem1&2; L1-3, loop1-3. A: linear sequential arrangement of the pseudoknot-forming sequence elements. Residues involved in the formation of S1 and S2 are represented as black and gray squares respectively. Residues in the single-stranded loop region are represented as unfilled circles. B: Schematic representations of folded pseudoknots. Left: with a non-zero L3 sequence; Middle: in the absence of L3, S1 and S2 can stack coaxially to form a quasi-continuous double helix. L1 and L2 locate on the same side of the double helix, with L1 crossing the major groove of S2 and L2 crossing the minor groove of S1; Right: The autoregulatory pseudoknot within the gene 32 mRNA of bacteriophages T2/T6 is shown as a representative of the CPK-1 (common pseudoknot motif-1) family of structurally related pseudoknots.

stimulating pseudoknots in animal viruses known or expected to use these recoding mechanisms (6). Using an in-house developed computer program capable of detecting all possible pseudoknots within the full-length genomic mRNAs (7), we found that ~85% of the recoding sites were associated with shortly downstream pseudoknots (established or putative). While these pseudoknots sample a wide range of stem and loop lengths, ~40% of the pseudoknots conform to a structurally related pseudoknot family known as CPK-1 (common pseudoknot motif 1) (8,9). A typical CPK-1

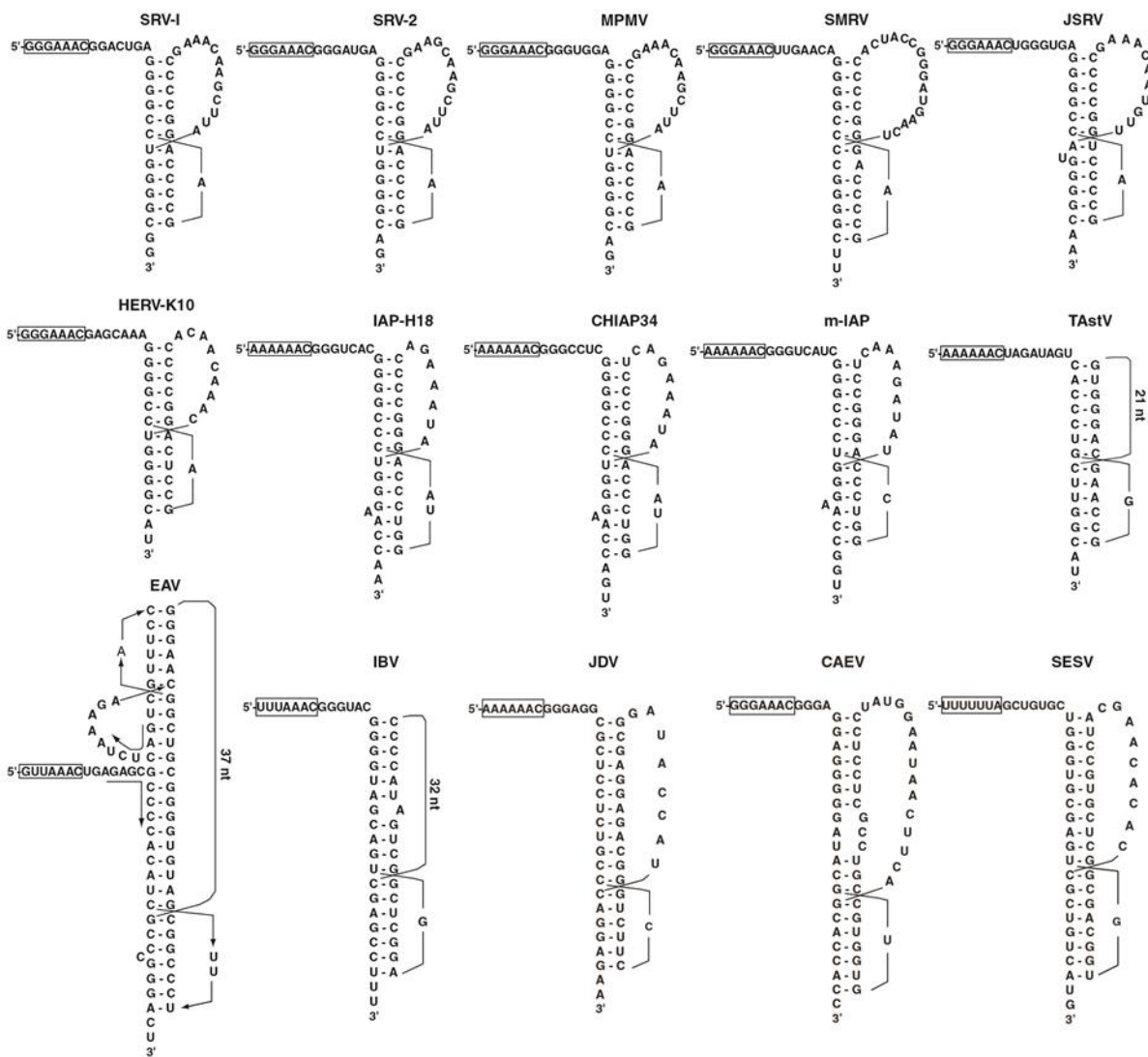
pseudoknot has a S2 of 6-7 base pairs and a very short L1 of 1-2 nucleotides; L3 is absent therefore the two helical stems S1 and S2 can stack coaxially to form a quasicontinuous helix (Figure 1B). It was noticed previously that many naturally occurring pseudoknots belonged to the CPK-1 family (8,9). With results from the systematic studies on viral frameshift and readthrough stimulating pseudoknots (6) (Figure 2, categories A and B), it becomes more evident that CPK1 pseudoknots occur in high abundance as -1 ribosomal frameshifting and readthrough stimulators in RNA viruses. CPK1 pseudoknots are also involved in the autogenous translational regulation of the bacteriophages T2/T6 gene 32 mRNAs (10-12) and the *E. coli* ribosomal protein S15 mRNA (13) (Figure 2, category C). These pseudoknots serve as a binding site for the protein products of the mRNAs, leading to translational autoregulation. Potential CPK-1 pseudoknots are also present in the internal ribosome entry site (IRES) elements of the cricket paralysis virus (CrPV) and CrPV-like viruses such as Plautia stall intestine virus (PSIV), *Drosophila C* virus (DCV), *Rhopalosiphum padi* virus (RhPV), himetobi P virus (HiPV), *Triatoma* virus (TrV), and black queen-cell virus (BQCV) (14) (Figure 2, category D). These pseudoknots may play a role in the IRES mediated methionine-independent initiation of translation. A few pseudoknots within the *E. coli* tmRNA may also belong to the CPK-1 family (Figure 2, category E). It is also interesting to note that some of the reported SELEX (Systematic Evolution of Ligands by Exponential Enrichment)-derived RNA aptamers have pseudoknot-forming sequences that may conform to the CPK-1 family, including aptamers that bind to biotin (15) and human nerve growth factor (NGF) (16) (Figure 2, the first and second pseudoknots in category E). Some of the HIV-1 reverse transcriptase (RT) binding pseudoknots have some CPK-1 features such as a stem2 of 6 base pairs, a loop1 of 2 nucleotides, and absence of loop3; however, the two stems are bent presumably due to the constraints

Figure 2: Pseudoknots known or expected to adopt a CPK1 structure. A) pseudoknots that stimulate stop codon readthrough in RNA viruses. The 0 frame stop condons are boxed. Mo-MuLV: Moloney murine leukemia virus; F-MuLV: Friend murine leukemia virus; MSV: murine sarcoma virus; BaEV: baboon endogenous virus; FeLV: feline leukemia virus; GaLV: gibbon ape leukemia virus; SNV: spleen necrosis virus. B) pseudoknots that stimulate -1 frameshifting in RNA viruses. The slippery sequences are boxed. Potential tandem pseudoknots are found in EAV and HIV-1. HCV-229E has an elaborated pseudoknot with an extra stem (stem3). SRV: simian retrovirus; MPMV: Mason-Pfizer monkey virus; SMRV: squirrel monkey retrovirus; JSRV: jaagsiekte sheep retrovirus; HERV: human endogenous retrovirus; IAP-H18, CHIAP34, and m-IAP: three intracisternal A particle (IAP) genetic elements from Syrian hamster, Chinese hamster, and mouse respectively; TAsV: turkey astrovirus; EAV: equine arteritis virus; IBV: avian infectious bronchitis coronavirus; JDV: Jembrana disease virus; CAEV: Caprine Arthritis Encephalitis virus; SESV: Seal louse virus; SA-OMVV: South African Ovine Maedi Visna virus; ScV: *Saccharomyces cerevisiae* virus; MHV: Murine hepatitis coronavirus; BEV: Berne Virus; HCoV: Human coronavirus; HIV: human immunodeficiency virus; HTLV: human T-cell leukemia virus; STLV: simian T-cell leukemia virus; PLV: puma lentivirus; FIV: feline immunodeficiency virus. C) pseudoknots that are involved in autogenous translational regulation in the gene 32 mRNA of bacteriophages T2/T6 and the *E. coli* ribosomal protein S15 mRNA. D) pseudoknots that are present in the IRES elements of cricket paralysis-like viruses, including the cricket paralysis virus (CrPV), *Drosophila C* virus (DCV), *Rhopalosiphum padi* virus (RhPV), himetobi P virus (HiPV), *Triatoma* virus (TrV), black queen-cell virus (BQCV), and *Plautia stall* intestine virus (PSIV). E) pseudoknots in the tmRNA of *E. coli*. F) SELEX pseudoknots that bind to biotin, human nerve growth factor (NGF), and HIV-1 reverse transcriptase (RT) respectively. The RT-binding pseudoknot has a bent conformation, probably due to constraints imposed by the short 3-nt loop2.

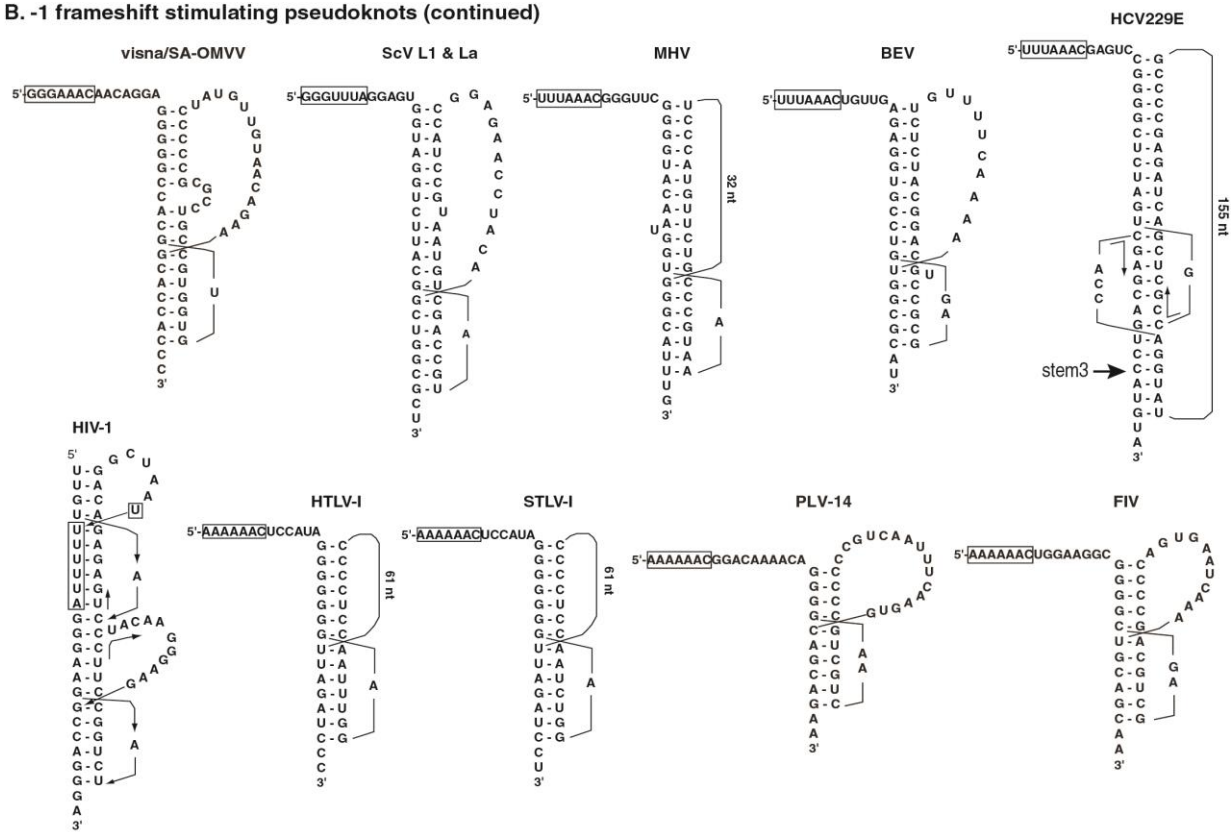
A. readthrough stimulating pseudoknots



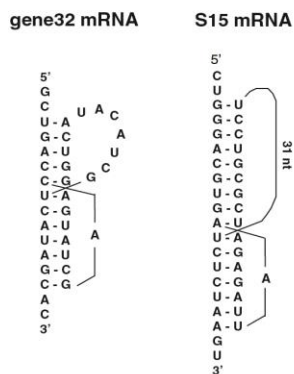
B. -1 frameshift stimulating pseudoknots



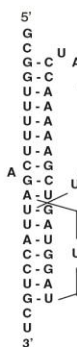
B. -1 frameshift stimulating pseudoknots (continued)



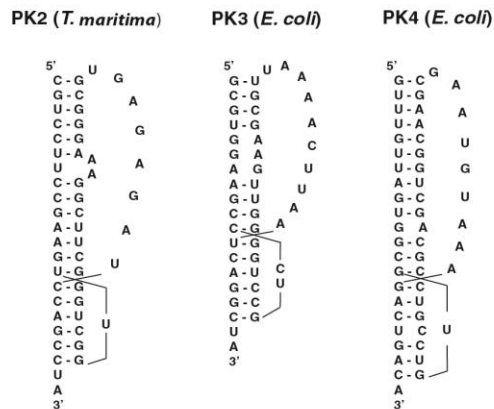
C. translational autoregulatory pseudoknots



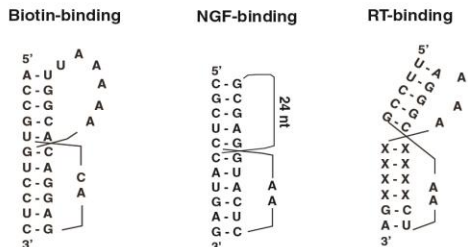
D. CrPV IRES pseudoknots



E. tmRNA pseudoknots



F. SELEX pseudoknots



imposed by a short loop2 of three nucleotides (17) (Figure 2, the last pseudoknot in category E). The RT-binding pseudoknot is therefore not a CPK-1 pseudoknot. The bent conformation of the pseudoknot may be critical for specific binding to the reverse transcriptase.

Structures of CPK-1 pseudoknots

Structures for several prototypic CPK-1 pseudoknots have been determined by NMR or X-ray crystallography, including the autoregulatory pseudoknot within the gene 32 mRNA of bacteriophages T2/T6 (8,9), the frameshift stimulating pseudoknot at the *gag-pro* junction of simian retrovirus-1 (SRV-1, the determined structure is for a functional mutant) (18,19), the readthrough stimulating pseudoknot at the *gag-pol* junction of murine leukaemia virus (MuLV) (20), and the SELEX biotin-binding pseudoknot (21) (Figure 3). There is no sequence homology among these pseudoknots, but they all have the following common structural features:

- 1) two coaxially stacked A-form stems;
- 2) a L1 with 1-2 nucleotides crossing the major groove of a S2 with 6-7 basepairs;
- 3) placement of the L1 bases inside the major groove of S2;
- 4) a major groove base triple between L1 and S2;
- 5) most importantly, the phosphodiester backbones of the interacting RNA strands in the regions defined by L1, S2, and S1-S2 junction assume similar structures in all four pseudoknots (Figure 3). In the high-resolution (1.3 Å) crystal structure of the biotin-binding pseudoknot (Figure 3C), there is a large electron density gap between the first and second nucleotides of L2, indicating that the RNA sequence is cut at this point (22). Therefore, the structure actually contains an intermolecular RNA-RNA binary complex. The original authors believed that the pseudoknot-forming RNA molecule was degraded into two pieces before crystallization took place, i.e. the intermolecular RNA-RNA binary complex also exists in solution as a stable structure.

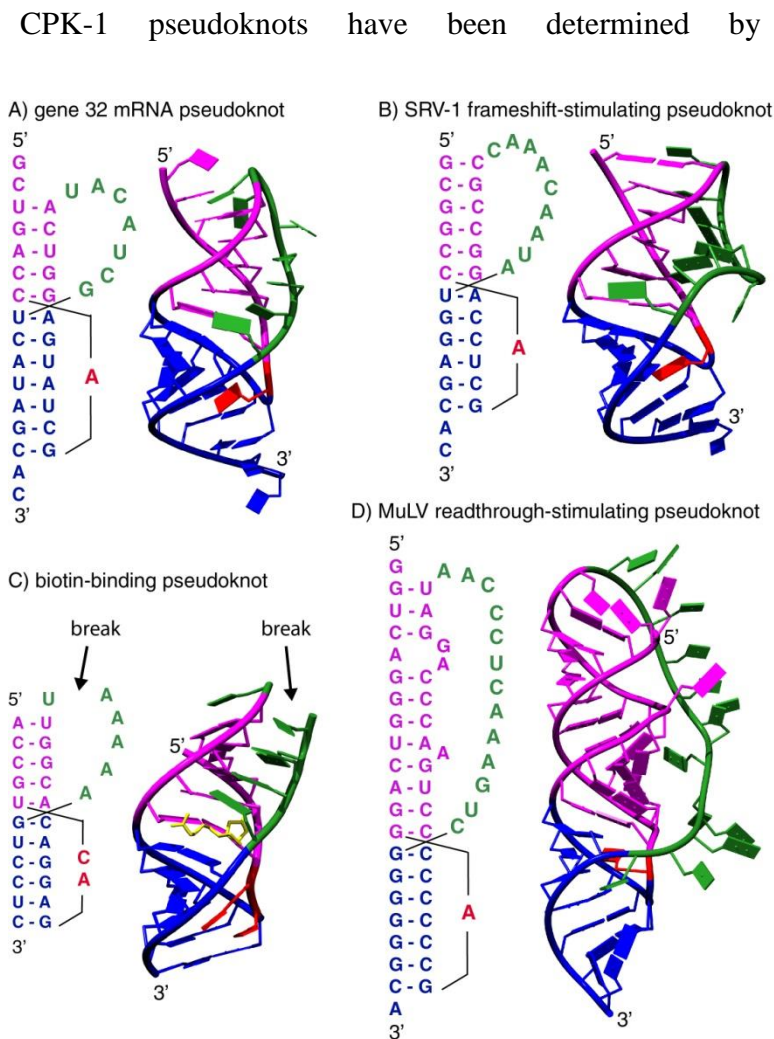


Figure 3: Secondary and tertiary structures of four CPK-1 pseudoknots. Stem 1, stem 2, loop1, and loop2 of the pseudoknots are colored differently in magenta, blue, red, and green respectively. A) the autoregulatory pseudoknot within the gene 32 mRNA of bacteriophages T2/T6 (PDB code: 2TPK). B) a frameshift-stimulating pseudoknot at the *gag-pro* junction of SRV-1 (PDB code: 1E95). Its sequence is different from the wild-type sequence (Figure 2B) at

3. RESULTS AND DISCUSSION

RNA-RNA interactions mediated by CPK-1-like structures

Results from the study on the biotin-binding pseudoknot show that the common structural features originally observed in compact CPK-1 pseudoknots can also mediate stable intermolecular RNA-RNA interactions. In naturally occurring systems, such an interaction may exist between the RNA-1 and the RNA-2 of the Dianthoviruses, including the Red clover necrotic mosaic virus (RCNMV), the Sweet clover necrotic mosaic virus (SCNMV) and the Carnation ringspot virus (CRSV) (Figure 4A). The trans-activator (TA) sequence located within the RNA-2 of these viruses harbors a simple hairpin. Six or seven nucleotides at the 3'-side of the TA loop may form complementary base pairs with the TA binding sequence (TABS) of the subgenomic promoter on RNA-1. NMR and structural modeling were used to characterize the intermolecular interaction in RCNMV. It was believed that the bimolecular complex featured two stacked helical stems and was structurally similar to the SRV-1 frameshift-stimulating pseudoknot (23). Conceivably, the common structural features of CPK-1 pseudoknots could also be able to mediate long-distance intramolecular RNA-RNA interactions between sequences separated by hundreds or thousands of nucleotides. In the frameshift stimulating pseudoknot in HCV229E (Figure 2, category B), L2 has 164 nt, the 3'-sequence of which forms an extra helix (stem3), resulting in an elaborated pseudoknot (24). The 155 nt between stem1 and stem3 may also harbour other secondary structures. This kind of pseudoknot may better be considered as a case of long-distance intramolecular RNA-RNA interactions instead of a compact CPK-1 pseudoknot. Theoretically speaking, if the number of nucleotides in L2 is much bigger than the number of basepairs in S1 (typically less than 15 bp in naturally occurring pseudoknots), there would be little constraint for the L2 nucleotides to position along the minor groove of S1 as seen in a compact pseudoknot.

Conserved structural features in different exhibitions

The CPK-1 conserved tertiary RNA structure is formed by complementary base-pairing between two intramolecular or intermolecular RNA strands. One of the RNA strands has a hairpin with 7-9 nucleotides in its loop. The 3'-sequence of this loop forms 6-7 complementary base-pairs with the other RNA strand, leaving only 1-2 unpaired loop nucleotides at the 5'-end (forming L1 in a pseudoknot configuration). The two stems (S1 and S2) have the potential to stack coaxially to form a quasicontinuous A-form helix. L1 crosses the major groove of S2. The base(s) of L1 is embedded in the major groove of S2. A major groove base triple between L1 and S2 is observed in four structures shown in Figure 3 (8,9,18-21), and therefore may represent a common structural feature of this kind of conserved RNA structures. Due to the natural twist of the phosphodiester backbones of an A-form RNA helix, the distance between two phosphorus atoms from opposing strands across the major groove reaches a minimum when the helical stem has 6-7 base pairs. This distance can be

bridged by a minimal number of 1 or 2 nucleotides that cross the major groove. The conserved tertiary RNA structure elegantly exploits these intrinsic structural features of an A-form RNA helix to achieve a maximal degree of base pairing and base stacking interactions. The base(s) of the major groove crossing nucleotide(s) may also interact with the stem (such as forming a base triple) to achieve increased stability. In the conserved RNA structure, the phosphodiester backbones of the interacting RNA strands adopt a distinctive folding topology that is sequence independent. Different sequences can assume similar overall structure. The strict requirement for the lengths of stem S2 (6-7 base pairs) and loop L1 (1-2 nucleotides) is dictated by the exploitation of A-form helical structural features. Adding nucleotides to or removing nucleotides from S2 and L1 would change the distinctive folding (and concomitantly the stability) of the structure. Conceivably, the conserved RNA structure can exhibit itself in three different forms, including compact CPK-1 pseudoknots, long-distance intramolecular RNA-RNA interactions, and intermolecular RNA-RNA interactions. In the form of compact CPK-1 pseudoknots, S1 and L2 are variable in terms of their lengths and compositions. The only restriction is that the length of L2 should be long enough to connect S1 and S2 without compromising their coaxial stacking. If L2 is too short, it may impose constraints on the two stems and induce bending at the helical junction. This scenario was observed in the SELEX RT-binding pseudoknot (Figure 2, category E). In the form of long-distance intramolecular interactions, the structure can be viewed as a CPK-1 pseudoknot with an extraordinary long L2.

Method of Detection

To detect potential presence of the conserved RNA structure (in any of the three forms described above), we used an in-house developed general purpose pseudoknot searching program (7). Figure 1A shows a linear arrangement of the sequence elements (S1, S2, L1, L2, and optionally L3) of a typical H-type pseudoknot. To detect potential pseudoknots in a given RNA sequence, the computer program scans the sequence and tests all possible combinations of stem and loop lengths within certain ranges (can be set by the user) to see whether both helical stems (S1 and S2) can form simultaneously. To detect compact CPK1 pseudoknots, the ranges for S2, L1 and L3 are set to 6-7 bp, 1-2 nt, and 0 nt respectively, the upper limit of L2 is set to a relatively small arbitrary number (e.g. 100 nt). To detect long-distance intramolecular interactions, the ranges for S2, L1 and L3 are set as in the case of detecting compact CPK1 pseudoknots, with the upper limit of L2 being increased to certain value according to the length of the given RNA sequence (the difference between the number of nucleotides in the given RNA sequence and the total number of nucleotides in S1, S2, and L1). The same settings can be used to detect intermolecular RNA-RNA interactions, except that the sequences of the two RNA molecules need to be merged into a single sequence as input and the detected structures need to be analyzed to keep only those intermolecular structures as outputs. Using the above described detection method, we have performed a search for potential

long-distance intramolecular interactions mediated by the conserved RNA structure in the HIV-1 genomic mRNAs (4254 different strains), with the following settings: S1=6-20 bp, S2=6-7 bp, L1=1-2 nt, L3=0, and L2=100-8000 nt. One possible case of such long-distance interaction (across more than one thousand intervening nucleotides) was detected (Figure 4B. Details will be communicated elsewhere). showing the usefulness of our program in the identification of potential hairpin sealer motif mediated long-distance intramolecular RNA-RNA interactions. Of course, any putative case (including most of those shown in Figure 2) needs to be confirmed by experimental methods.

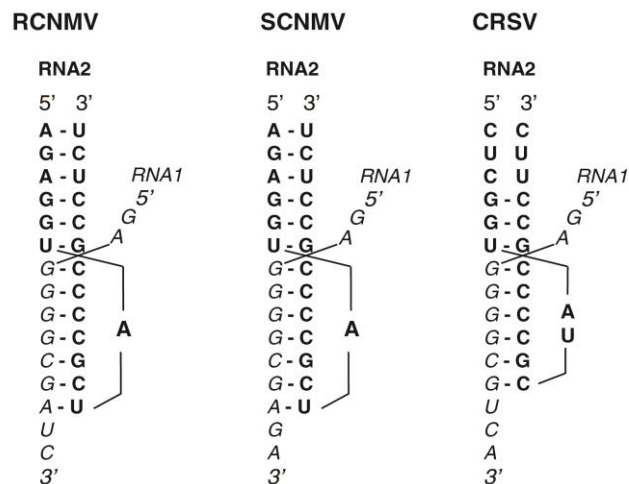
Biological Functions

RNA molecules harbouring CPK-1 and CPK-1-like structures participate in many different biological functions. As shown in categories A-D in Figure 2, these functions include: -1 ribosomal frameshifting, stop codon readthrough, autogenous translational regulation, and IRES mediated initiation of translation. The two NGF-binding and biotin-binding SELEX pseudoknots further suggest that the structures in the form of compact CPK-1 pseudoknots can provide specific binding sites not only for proteins, but also for small molecules. Although all of the CPK-1 pseudoknots share common features defined by the conserved RNA structure, the biological functions of the pseudoknots are interchangeable. For example, it was found that the autoregulatory pseudoknot in bacteriophage T2/T6 gene 32 mRNA was unable to function as a -1 frameshift stimulator (25). Different frameshift stimulating pseudoknots may also differ in their frameshift stimulating efficiencies. The pseudoknots in Figure 2 show a fair degree of variations in the lengths and compositions of S1 and L2. Some of the pseudoknots may also harbour additional structural features (such as seen in the EAV and HCV229E pseudoknots) which would make the pseudoknots even more versatile. Conceivably, the common features defined by the conserved RNA structure may primarily serve a structural role for maintaining a stable and compact pseudoknot scaffold while the more variable parts of the pseudoknots (including S1, L2, and possible interactions between them) contribute more to the diversity and fine-tuning of the biological functions. Of course, sequence specific interactions between L1 and S2 may also be functionally important in certain cases. For example, it was found that protonation of the L1 adenosine in the readthrough stimulating pseudoknot of MuLV (Figures 2A and 3D) promoted the formation of a L1-S2 base-triple, which is required for readthrough stimulating activity of the pseudoknot (20).

So far, most all of the established or putative RNA hairpin sealer motif mediated structures belong to the category of compact CPK-1 pseudoknots (Figure 2). Formation of these pseudoknots only requires a localized, relatively short sequence, which makes the detection of these pseudoknots relatively easy. It is more difficult to discover hairpin sealer motif mediated structures belonging to the other two categories (long-distance intramolecular interactions and intermolecular interactions). The HCV229E structure shown in Figure 2 is a putative case of long-distance intramolecular interactions that stimulates efficient -1 ribosomal frameshifting. The RNA1-RNA2 interactions in the Dianthoviruses (Figure 4A) are putative cases of long-distance intermolecular interactions mediated by the hairpin sealer motif. The interactions transactivate subgenomic RNA synthesis.

It is well established that long-distance intramolecular RNA-RNA interactions (crossing hundreds or thousands of intervening nucleotides) (26-39) and intermolecular RNA-RNA interactions (40-48) can play critical regulatory roles in a variety of biological processes including translation, replication, subgenomic mRNA transcription, mRNA splicing, etc. Although the number of potential cases of hairpin sealer motif mediated long-distance intramolecular interactions and intermolecular interactions are small at this time, it is fully expected that more cases with various functions will be uncovered and confirmed in the future.

A) Intermolecular interaction



B) Long-distance intramolecular interaction

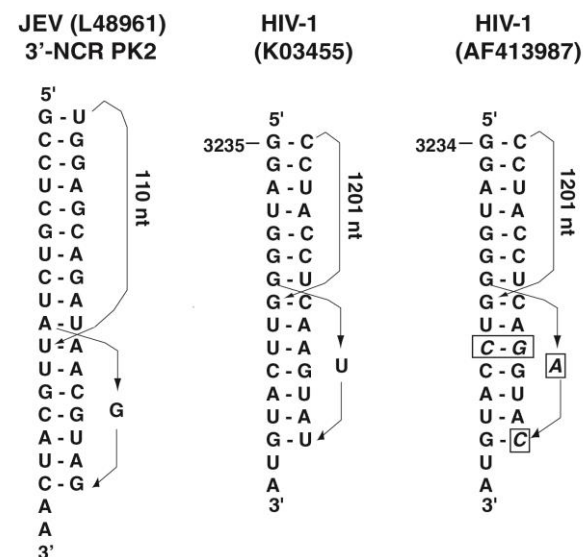


Figure 4: Potential cases of hairpin sealer motif mediated intermolecular and long-distance intramolecular interactions. A) Intermolecular interactions between RNA-1 (italic) and RNA2 of three Dianthoviruses: Red clover necrotic mosaic virus (RCNMV), Sweet clover necrotic mosaic virus (SCNMV) and Carnation ringspot virus (CRSV). B) Long-distance intramolecular interactions in two HIV-1 strains, including the reference strain of HXB2 (K03455). Sequence variations in the strain AF413987 are highlighted with boxes.

An antisense strategy based on the CPK-1 like conserved structure

The ability of the CPK-1-like structures to mediate stable intermolecular RNA-RNA interactions can potentially be exploited to develop a structure-based antisense strategy. The principle is well demonstrated by the proof of concept structure of intermolecular RNA-RNA interaction in Figure 3C. In this structure, the target RNA contains a stem-loop with 8 nt ($C_6AGAGGAC_{13}$) in the loop. This stem-loop is targeted by an antisense oligonucleotide ($A_{20}AAAAGUCCUC_{30}$) that contains a sequence (underlined) complementary to all but 2 nt of the loop. Base pairing interaction between the hairpin and the antisense oligonucleotide results in the formation of a CPK-1-like structure mediating intermolecular interaction. Theoretically, any RNA hairpin with 7-9 nucleotides in its loop can be exploited as a target. The antisense oligonucleotide should contain extra residues flanking the complementary sequence at both the 5'- and 3'- sides. A 2-3 nt overhang at the 3'-side should have a stabilizing effect on the newly formed helical stem. The extra nucleotides at the 5'-side may stabilize the helical junction. These nucleotides may also be designed in such a way that they could form minor groove base triples with the stem region of the target hairpin. These extra interactions may increase the binding affinity and specificity of the antisense strategy. Because DNA-RNA and locked nucleic acid (LNA)-RNA hybrid duplexes adopt A-form helical structures (57-59), DNA or LNA oligonucleotides may also be able to form CPK-1-like structures that mediate intermolecular interaction with the target RNA. These types of oligonucleotides may have some advantages over RNA oligonucleotides in terms of cost and/or stability as antisense oligonucleotides. Previously, a pseudo-half-knot motif was proposed in which an antisense oligonucleotide binds asymmetrically to the loop region of an RNA hairpin. So far, there is no structure determination for any pseudo-half-knot motif mediated structure. Presumably, the newly formed helix stacks coaxially with the stem region of the hairpin. It was shown that pseudo-half-knot-forming oligonucleotides complementary to HIV TAR element altered the TAR structure and disrupted tat-TAR interaction (60). In another study, a pseudo-half-knot-forming oligonucleotide was used as a cofactor to control the activity of a designed ribozyme (61). According to the definition of pseudo-half-knot, a CPK-1-like structure mediating intermolecular RNA-RNA interaction is actually a special case of pseudo-half-knot in which a structurally conserved and naturally preferred interaction is formed.

4. CONCLUSION

We propose the existence of a conserved RNA structure that can manifest itself in various forms, including compact pseudoknots, long-distance intramolecular interactions, and intermolecular interactions. These structures have a wide-spread occurrence RNA in natural RNAs and assume diverse biological functions. Revelation of these structures should have positive impacts on many fronts of RNA-related studies, such as prediction of RNA tertiary structures from sequence data,

RNA engineering, and targeting regulatory RNA structures with antisense strategy.

CONFLICT OF INTEREST

The authors have no conflict of interest.

ACKNOWLEDGEMENT

This work was supported by the start-up fund and a seed grant to Z. D. from Southern Illinois University Carbondale

REFERENCES

1. Pleij, C.W., Rietveld, K. and Bosch, L. (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.*, 13, 1717-1731.
2. Pleij, C.W. (1990) Pseudoknots: a new motif in the RNA game. *Trends in Biochemical Science*, 15, 143-147.
3. Gesteland, R.F. and Atkins, J.F. (1996) Recoding: dynamic reprogramming of translation. *Annu Rev Biochem*, 65, 741-768.
4. Gesteland, R.F., Weiss, R.B. and Atkins, J.F. (1992) Recoding: reprogrammed genetic decoding. *Science*, 257, 1640-1641.
5. Baranov, p.v., Gesteland, R.F. and Atkins, J.F. (2002) Recoding: translational bifurcations in gene expression. *Gene*, 286, 187-201.
6. Huang, X., Cheng, Q. and Du, Z. (2013) A genome-wide analysis of RNA pseudoknots that stimulate efficient -1 ribosomal frameshifting or readthrough in animal viruses. *Biomed Res Int*, 2013, 984028.
7. Huang, X., Du, Z., Cheng, J. and Cheng, Q. (2013) PKscan: a program to identify H-type RNA pseudoknots in any RNA sequence with unlimited length. *Bioinformatics*, 9, 440-442.
8. Du, Z., Giedroc, D.P. and Hoffman, D.W. (1996) Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots. *Biochemistry*, 35, 4187-4198.
9. Du, Z. and Hoffman, D. (1997) An NMR and mutational study of the pseudoknot within the gene 32 mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots. *Nucl. Acids. Res.*, 25, 1130-1135.
10. Qiu, H., Kaluarachchi, K., Du, Z., Hoffman, D.W. and Giedroc, D.P. (1996) Thermodynamics of folding of the RNA pseudoknot of the T4 gene 32 autoregulatory messenger RNA. *Biochemistry.*, 35, 4176-4186.
11. McPheeters, D.S., Stormo, G.D. and Gold, L. (1988) Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. *J Mol Biol.*, 201, 517-535.

12. Du, Z., Giedroc, D.P. and Hoffman, D.W. (1996) Structure of the Autoregulatory Pseudoknot within the Gene 32 Messenger RNA of Bacteriophages T2 and T6: A Model for a Possible Family of Structurally Related RNA Pseudoknots. *Biochemistry*, 35, 4187-4198.
13. Philippe, C., Benard, L., Portier, C., Westhof, E., Ehresmann, B. and Ehresmann, C. (1995) Molecular dissection of the pseudoknot governing the translational regulation of Escherichia coli ribosomal protein S15. *Nucleic Acids Res.*, 23, 18-28.
14. Kanamori, Y. and Nakashima, N. (2001) A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation. *RNA*, 7, 266-274.
15. Wilson, C., Nix, J. and Szostak, J. (1998) Functional Requirements for Specific Ligand Recognition by a Biotin-Binding RNA Pseudoknot. *Biochemistry*, 37, 14410-14419.
16. Hofstadler, S.A., Sannes-Lowery, K.A., Crooke, S.T., Ecker, D.J., Sasmor, H., Manalili, S. and Griffey, R.H. (1999) Multiplexed screening of neutral mass-tagged RNA targets against ligand libraries with electrospray ionization FTICR MS: a paradigm for high-throughput affinity screening. *Anal Chem*, 71, 3436-3440.
17. Swayze, E.E., Jefferson, E.A., Sannes-Lowery, K.A., Blyn, L.B., Risen, L.M., Arakawa, S., Osgood, S.A., Hofstadler, S.A. and Griffey, R.H. (2002) SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J Med Chem*, 45, 3816-3819.
18. Du, Z., Holland, J.A., Hansen, M.R., Giedroc, D.P. and Hoffman, D.W. (1997) Base-pairings within the RNA pseudoknot associated with the simian retrovirus-1 gag-pro frameshift site. *J Mol Biol.*, 270, 464-470.
19. Michiels, P.J., Versleijen, A.A., Verlaan, P.W., Pleij, C.W., Hilbers, C.W. and Heus, H.A. (2001) Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J Mol Biol.*, 310, 1109-1123.
20. Houck-Loomis, B., Durney, M.A., Salguero, C., Shankar, N., Nagle, J.M., Goff, S.P. and D'Souza, V.M. (2011) An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature*, 480, 561-564.
21. Nix, J., Sussman, D. and Wilson, C. (2000) The 1.3 angstrom crystal structure of a biotin-binding pseudoknot and the basis for RNA molecular recognition. *J Mol Biol*, 296, 1235-1244.
22. Nix, J., Sussman, D. and Wilson, C. (2000) The 1.3 Å crystal structure of a biotin-binding pseudoknot and the basis for RNA molecular recognition. *J Mol Biol*, 296, 1235-1244.
23. Guenther, R.H., Sit, T.L., Gracz, H.S., Dolan, M.A., Townsend, H.L., Liu, G., Newman, W.H., Agris, P.F. and Lommel, S.A. (2004) Structural characterization of an intermolecular RNA-RNA interaction involved in the transcription regulation element of a bipartite plant virus. *Nucleic Acids Res*, 32, 2819-2828.

24. Herold, J. and Siddell, S.G. (1993) An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res.*, 21, 5838-5842.
25. Wang, Y., Wills, N.M., Du, Z., Rangan, A., Atkins, J.F., Gesteland, R.F. and Hoffman, D.W. (2002) Comparative studies of frameshifting and nonframeshifting RNA pseudoknots: a mutational and NMR investigation of pseudoknots derived from the bacteriophage T2 gene 32 mRNA and the retroviral gag-pro frameshift site. *RNA*, 8, 981-996.
26. Barry, J.K. and Miller, W.A. (2002) A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc Natl Acad Sci U S A*, 99, 11133-11138.
27. Tajima, Y., Iwakawa, H.O., Kaido, M., Mise, K. and Okuno, T. (2011) A long-distance RNA-RNA interaction plays an important role in programmed -1 ribosomal frameshifting in the translation of p88 replicase protein of Red clover necrotic mosaic virus. *Virology*, 417, 169-178.
28. Miller, W.A. and White, K.A. (2006) Long-distance RNA-RNA interactions in plant virus gene expression and replication. *Annu Rev Phytopathol*, 44, 447-467.
29. Keel, A.Y., Jha, B.K. and Kieft, J.S. (2012) Structural architecture of an RNA that competitively inhibits RNase L. *RNA*, 18, 88-99.
30. Romero-Lopez, C., Barroso-Deljesus, A., Garcia-Sacristan, 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.
31. Mateos-Gomez, P.A., Morales, L., Zuniga, S., Enjuanes, L. and Sola, I. (2013) Long-distance RNA-RNA interactions in the coronavirus genome form high-order structures promoting discontinuous RNA synthesis during transcription. *J Virol*, 87, 177-186.
32. Basu, M. and Brinton, M.A. (2011) West Nile virus (WNV) genome RNAs with up to three adjacent mutations that disrupt long distance 5'-3' cyclization sequence basepairs are viable. *Virology*, 412, 220-232.
33. Chattopadhyay, M., Shi, K., Yuan, X. and Simon, A.E. (2011) Long-distance kissing loop interactions between a 3' proximal Y-shaped structure and apical loops of 5' hairpins enhance translation of Saguaro cactus virus. *Virology*, 417, 113-125.
34. Hu, B., Pillai-Nair, N. and Hemenway, C. (2007) Long-distance RNA-RNA interactions between terminal elements and the same subset of internal elements on the potato virus X genome mediate minus- and plus-strand RNA synthesis. *RNA*, 13, 267-280.
35. Kim, K.H. and Hemenway, C.L. (1999) Long-distance RNA-RNA interactions and conserved sequence elements affect potato virus X plus-strand RNA accumulation. *RNA*, 5, 636-645.

- 36.Lin, H.X. and White, K.A. (2004) A complex network of RNA-RNA interactions controls subgenomic mRNA transcription in a tombusvirus. *EMBO J*, 23, 3365-3374.
- 37.Moreno, J.L., Zuniga, S., Enjuanes, L. and Sola, I. (2008) Identification of a coronavirus transcription enhancer. *J Virol*, 82, 3882-3893.
- 38.van Himbergen, J., van Geffen, B. and van Duin, J. (1993) Translational control by a long range RNA-RNA interaction; a basepair substitution analysis. *Nucleic Acids Res*, 21, 1713-1717.
- 39.Zhang, G., Slowinski, V. and White, K.A. (1999) Subgenomic mRNA regulation by a distal RNA element in a (+)-strand RNA virus. *RNA*, 5, 550-561.
- 40.Ferrandon, D., Koch, I., Westhof, E. and Nusslein-Volhard, C. (1997) RNA-RNA interaction is required for the formation of specific bicoid mRNA 3' UTR-STAUFIN ribonucleoprotein particles. *EMBO J*, 16, 1751-1758.
- 41.Guo, P., Zhang, C., Chen, C., Garver, K. and Trottier, M. (1998) Inter-RNA interaction of phage phi29 pRNA to form a hexameric complex for viral DNA transportation. *Mol Cell*, 2, 149-155.
- 42.Kolb, F.A., Engdahl, H.M., Slagter-Jager, J.G., Ehresmann, B., Ehresmann, C., Westhof, E., Wagner, E.G. and Romby, P. (2000) Progression of a loop-loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA. *EMBO J*, 19, 5905-5915.
- 43.Ly, H., Xu, L., Rivera, M.A., Parslow, T.G. and Blackburn, E.H. (2003) A role for a novel 'trans-pseudoknot' RNA-RNA interaction in the functional dimerization of human telomerase. *Genes Dev*, 17, 1078-1083.
- 44.Mikkelsen, J.G., Lund, A.H., Duch, M. and Pedersen, F.S. (2000) Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination in vivo. *J Virol*, 74, 600-610.
- 45.Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C. and Marquet, R. (1996) A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A*, 93, 5572-5577.
- 46.Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C. and Marquet, R. (2002) In vitro evidence for a long range pseudoknot in the 5'-untranslated and matrix coding regions of HIV-1 genomic RNA. *J Biol Chem*, 277, 5995-6004.
- 47.Skripkin, E., Isel, C., Marquet, R., Ehresmann, B. and Ehresmann, C. (1996) Psoralen crosslinking between human immunodeficiency virus type 1 RNA and primer tRNA³(Lys). *Nucleic Acids Res*, 24, 509-514.
- 48.Sun, J.S. and Manley, J.L. (1995) A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev*, 9, 843-854.

49. Deng, J., Xiong, Y., Sudarsanakumar, C., Shi, K. and Sundaralingam, M. (2001) Crystal structures of two forms of a 14-mer RNA/DNA chimer duplex with double UU bulges: a novel intramolecular U*(A x U) base triple. *RNA*, 7, 1425-1431.
50. Han, G.W. (2001) Direct-methods determination of an RNA/DNA hybrid decamer at 1.15 Å resolution. *Acta Crystallogr D Biol Crystallogr*, 57, 213-218.
51. Nielsen, K.E., Rasmussen, J., Kumar, R., Wengel, J., Jacobsen, J.P. and Petersen, M. (2004) NMR studies of fully modified locked nucleic acid (LNA) hybrids: solution structure of an LNA:RNA hybrid and characterization of an LNA:DNA hybrid. *Bioconjug Chem*, 15, 449-457.
52. Ecker, D.J., Vickers, T.A., Bruice, T.W., Freier, S.M., Jenison, R.D., Manoharan, M. and Zounes, M. (1992) Pseudo--half-knot formation with RNA. *Science*, 257, 958-961.
53. Komatsu, Y., Yamashita, S., Kazama, N., Nobuoka, K. and Ohtsuka, E. (2000) Construction of new ribozymes requiring short regulator oligonucleotides as a cofactor. *J Mol Biol*, 299, 1231-1243.