Identification of two novel putative satellite RNAs with hammerhead structures in the virome of French and Spanish carrot samples

B M Babalola, Deborah Schönegger, Chantal Faure, Armelle Marais, A. Fraile, F. Garcia-Arenal, Thierry T. Candresse

Identification of two novel putative satellite RNAs with hammerhead structures in the virome of French and Spanish carrot samples

B.M. Babalola[1], **D. Schönegger**[2], **C. Faure**[2], **A. Marais**[2$], **A. Fraile**[1], **F. Garcia-Arenal**[1] and **T. Candresse**[2*]

[1]Centro de Biotecnología y Genómica de Plantas UPM-INIA, E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, Universidad Politécnica de Madrid, 28223, Pozuelo de Alarcón, Madrid, Spain.

[2]Univ. Bordeaux, INRAE, UMR BFP, CS 20032, 33882 Villenave d'Ornon CEDEX, France

*Corresponding author: thierry.candresse@inrae.fr ORCID 0000-0001-9757-1835

$ ORCID 0000-0003-2482-1543

Running title: Novel hammerhead satellites in carrot virome

Key words: satellite, high-throughput sequencing, ribozyme, Polerovirus

This version of the article has been accepted for publication after peer review. It is not the Version of Record and does not reflect post-acceptance improvements. The Version of Record is available online at: http://dx.doi.org/ 10.1007/s00705-022-05538-z

Abstract

Carrot virome analysis using high-throughput sequencing revealed the presence of two RNA molecules with properties of satellite RNAs and homologous to the satellite RNA of cereal yellow dwarf virus-RPV (CYDV-RPV). Satellite 1 is 298 nt long, while Satellite 2 is 368 nt long. Their positive and negative genome strands contain hammerhead ribozymes similar to those found in other self-cleaving satellite RNAs. While both satellites were detected in Spanish carrot populations, only satellite 2 was found in French carrot populations. The most likely helper virus for these two satellites is carrot red leaf virus (CtRLV), which similar to CYDV-RPV is a *Polerovirus*.

Satellites RNAs (satRNAs) are subviral agents that are dependent on their helper virus (HV) for critical functions such as replication, encapsidation or transmission [1-3]. Contrary to defective RNAs, satRNAs have no sequence homology with their HV. Some of them are capable of modifying, positively or negatively, the disease symptoms induced in host plants by the HV while others do not seem to substantially affect their HV or the symptoms it causes [1-3]. For example, the satRNA of tobacco ringspot virus (TRSV satRNA) reduces the symptoms of TRSV, while that of arabis mosaic virus (ArMV satRNA) intensifies ArMV symptoms [1].

Several groups of satRNAs have been distinguished based on genome size and properties [1-2]. While some satRNAs have large genomes (in the 1 kb and above range) that harbour open reading frames (ORFs), smaller non-coding linear satellite RNAs (sl-SatRNA) with genomes less than 700 nt have been reported with HVs in the families *Tombusviridae* and *Bromoviridae*. Another class is represented by small circular non-coding satellite RNAs (sc-SatRNA) with genomes of about 300-350 nt and which are mostly associated with helper viruses in the families *Secoviridae* (e.g arabis mosaic virus small satellite RNA, ArMV satRNA) and *Solemoviridae* such as the satellites associated with several members of the *Sobemovirus* genus (e.g rice yellow mottle virus satellite, RYMV-SatRNA) [4]. The sc-SatRNAs are known to replicate by a rolling circle mechanism and to generate unit-length genomes through self-cleavage [5, 6]. They show a high degree of secondary structure [1] and have in their positive strand genome conserved sequences that correspond to self-cleaving “hammerhead” ribozymes [6, 7]. Depending on the satellite, their minus strand genome either does not encode a ribozyme, or encodes a hairpin or a hammerhead ribozyme [8]. So far, the only small satellite with hammerhead ribozymes on both genome strands and with a HV in the *Polerovirus* genus is cereal yellow dwarf virus-RPV satellite RNA (CYDV-RPV satRNA) [8, 9]. While initially described as a satellite of the *Luteovirus* barley yellow dwarf

virus-RPV (BYDV-RPV), a later reclassification and splitting of the *Luteovirus* genus led to a renaming of the HV as cereal yellow dwarf virus-RPV (CYDV-RPV) in the genus *Polerovirus* [10]. The replication of CYDV-RPV satRNA and the processing of the multimeric forms to monomeric copies by the hammerhead ribozymes encoded on its plus and minus genome strands have been extensively studied [9, 11-14]. These studies have in particular demonstrated that the replication of the satellite is dependent on the HV [11], identified key residues in the plus and minus strand ribozymes [13] as well as tentative replicase recognition sites on both satellite strands and a putative encapsidation signal [14]. Thus, a replication model was proposed in which alternative conformations of the molecule favor ribozyme activity or, on the contrary efficient CYDV-RPV satRNA RNA replication through ribozyme activity inhibition [12-14]. CYDV-RPV satRNA has also been shown to reduce accumulation of its HV and attenuate its symptoms [15] and to be able to be assisted by beet western yellows virus (another *Polerovir*us) in tobacco protoplasts as well as in *Capsella bursa-pastoris* plants [16].

The present work is focused on the characterization of two new putative hammerhead-containing sc-SatRNAs (Hhsats) discovered in the virome of carrot red leaf virus-infected Spanish and French carrots (*Daucus carota* L.).

In June 2021, five cultivated carrot (*Daucus carota spp sativa*) fields and one wild carrot (*Daucus carota spp carota*) population were sampled near Segovia (Central Spain). From each population, fifty plants were sampled (irrespective of whether they showed symptoms of viral infection such as leaf reddening or yellowing) and assembled in a pool. Double stranded RNAs were extracted from each pool [17], converted to complementary DNA, and finally sequenced (2x125 nt paired reads, Illumina Hiseq2500). Following reads cleanup, contigs were assembled *de novo* using CLC Genomics Workbench v22.0 and annotated by BlastN and BlastX analysis against the Virus

section of the GenBank RefSeq database. Two groups of contigs of 293 to 368 nt with distant homology to CYDV-RPV satRNA (M63666) were identified in the virome of several pools. Extension of these contigs by rounds of mapping of residual reads allowed us to assemble partial multimers of two different putative small satellite RNA molecules from which unit-length monomeric sequences were derived. To confirm the presence of the two satellites in sampled carrots, a two-step RT-PCR assay with specific detection primers designed using the genome sequence of each of the satellites (Sat1_F: 5' ACAGAAAACCACCCGAGTAA 3' & Sat1_R: 5' TAACCACATGGGAGTCATCCT 3', and Sat2_F: 5' CCACCACACTCGTTTTGTG 3' & Sat2_R: 5' TCCACTTCTTCCTCGATTGAG 3') were used. Both primer pairs generate a 258 nt amplicon. Sat1 was present in all six sampled carrot populations whereas Sat2 was found only in four cultivated carrot populations. Sanger sequencing of the amplicons for the two satellites showed them to be identical to the respective sequences obtained by assembly of Illumina reads. In addition, the presence of circular versions of these two molecules was verified by RT-PCR using pairs of specific, divergent primers: Sat1-R1 5'-GTCTCCTCACTTCAAAGAGTG-3' and Sat1-F1 5'-GCTTTACGTGTCTGTCATCAA-3' and Sat2-R1 5'-TACCTCGACTGATGAGTTCAA-3' and Sat2-F1 5'-GCACCTCGAGACACCTTTCCT-3'.

The smallest of the two molecules, which is referred to as Sat1 is 298 nt long, while the longest one, referred to as Sat2, is 368 nt long. These sequences have been deposited in GenBank under accession numbers OM962993 and OM962994, respectively. The two molecules are 70.7% identical while they respectively share 84.6% (Sat1) and 69.4% (Sat2) identity with CYDV-RPV satRNA. In particular, they share two regions of high homology corresponding with the conserved hammerhead ribozymes on the plus and minus polarity of their genome (corresponding to positions 1-77 and 206-260 of the Sat2 molecule). In these regions, Sat1 and Sat2 are respectively 96% and

92.3% identical and share respectively 93.3-96% and 87.3-90.9% identity with the corresponding regions of CYDV-RPV satRNA (Figures 1A and 1B). Given this high homology level, the proposed folding for the plus and minus strand ribozymes are similar to those determined by Miller et al. [9, 12]. In both hammerheads, nucleotide differences in hairpin regions with respect to the hammerhead ribozymes of CYDV-RPV satRNA are systematically accompanied by compensatory changes on the other hairpin strand [bases 196 & 203, 194 &229 & and 195 & 228 for the minus strand ribozyme (Figure 1A) and bases 1 &16, 2 &15, 78 & 296) for the plus strand one (Figure 1B)], providing a strong evolutionary support for the secondary structure proposed for CYDV-RPV satRNA. All other nucleotide changes as compared to CYDV-RPV satRNA hammerheads occur in loop regions (position 222 for the minus strand ribozyme, position 21 for the plus strand one), with the exception of several mutations (positions 287-289 and 291-290 plus 84-85) that shorten the lateral stem 3 (naming according to [12]). In the same fashion, the double hammerhead structure that has been proposed [13] to be the active form for the processing of the plus strand multimers shows compensatory mutations so that the possibility to form stem I is fully conserved (not shown).

Remarkably the region of CYDV-RPV satRNA that is essential for replication and has been proposed to be a replicase binding site or an origin of replication (positions 245-310, [14], Figure 2A) is the most divergent region between the different satellites, showing both large length variation (respectively 66, 50 and 107 nt for CYDV-RPV satRNA, Sat1 and Sat2) and low homology (only 34% identity between Sat1 and Sat2 and respectively 68.1% and 43.9% identity between CYDV-RPV satRNA and Sat1 or Sat2) (Figure 2). Similarly, the CYDV-RPV satRNA 152-194 region, which has been suggested to be involved in encapsidation [14] (Figure 2A) is poorly conserved in the two carrot satellites, showing again both length variation (respectively 43,

35 and 31 nt for CYDV-RPV satRNA, Sat1 and Sat2) and sequence divergence (only 54.8% between Sat1 and Sat2 and respectively 68.6% and 58.1% between CYDV-RPV satRNA and Sat1 or Sat2) (Figure 2). Taken together, the secondary structure of CtRLV Sat1 and Sat2 reveal three kinds of functional elements present in CYDV-RPV satRNA, that is (i) self-functioning ribozymes; (ii) elements that control conformational change (RNA switches, e.g. L1 and L2a that facilitate replication); and (iii) cis-acting elements that interact with helper virus and/or host components, e.g., origins of replication and assembly [14].

As judged by the number of HTS reads corresponding to each molecule, the smaller Sat1 consistently showed much higher accumulation levels than the larger Sat2. For example, in the third cultivated carrot population, Sat1 represented 170,232 reads (70,155x coverage, 0.8% of total reads for the sample), while Sat2 represented only 2,218 reads (739x coverage, 0.01% of total reads), which converts to a 95-fold higher representation for Sat1. Given that the HTS data was generated from pools of plants, it is however not known whether the two Hhsats are jointly present in the same plants and therefore if this observed variation in representation reflects a true difference in accumulation in infected plants, a difference in prevalence of the two molecules in the sampled carrot populations or a combination of both factors.

Full genomes could be assembled from HTS sequencing reads for five carrot populations for Sat1 but could only be reconstructed for two populations for Sat2 due to its lower representation. The Sat1 genomic sequences obtained for the five Spanish populations are strictly identical. For Sat2, a single indel polymorphism is observed between the two populations, $_{123}$UUUU$_{126}$ > UUU. Analysis of HTS datasets for carrot populations collected in France in 2020 and processed in a similar fashion as the Spanish samples indicated the presence of Sat2, but not that of Sat1 in two populations collected in Aquitaine about 20 km apart. One of these French populations

corresponds to cultivated carrots, while the second corresponds to wild carrots. Given the 81x-193x average coverage of Sat2 in the HTS reads for these two populations, full-length genomic sequences could be reconstructed (GenBank OM962996). The two sequences are identical and show a single polymorphism as compared to the Spanish sequences, which affects the region that is also polymorphic between the two Spanish isolates ($_{123}$U**A**U$_{126}$). The variability between the carrot satellites found in France and Spain is therefore low, while there is no evidence so far for the presence of Sat1 in France despite having sampled to date 45 carrot populations representing a total of 2250 plants. Given the differential representation of Sat1 in Spanish samples, it seems unlikely that if present in French samples it could have been missed during these virome studies.

HhSats have been found associated with nepoviruses in the *Secoviridae* and sobemoviruses and poleroviruses in the *Solemoviridae*. Although HhSats have no sequence homology with their helpers, BlastN analysis of Sat1 and Sat2 showed significant homologies to CYDV-RPV satRNA which has a polerovirus HV and can be assisted by at least another polerovirus [16] but detected no homologies to other HhSats. A detailed analysis of the virome of the carrot samples in which Sat1 and Sat2 were identified shows that the most likely HV candidate is carrot red leaf virus (CtRLV), which similar to CYDV-RPV is a *Polerovirus*. In fact, CtRLV was the only *Solemoviridae* identified in these plants pools. The other viruses identified in the virome belong to genera not known to be associated with HhSats and a potential role of carrot torradovirus 1 (CaTV1), a *Torradovirus* in the *Secoviridae* family can be excluded as it was not identified in the virome of the French carrot samples in which Sat2 was identified. Consequently, CtRLV appears as the most likely HV for these two novel putative satellites and the names CtRLV satellites 1 and 2 are therefore proposed for the two molecules identified here. To our knowledge, these results represent the first identification of satellites RNAs associated with a polerovirus under field

148 conditions, since CYDV-RPV satRNA was first described in a laboratory isolate of its HV and
149 does not appear to have been observed again under field conditions. Since there were no specific
150 efforts to select for symptomatic or asymptomatic plants to assemble the pools, it is not possible
151 to know whether Sat1 or Sat2 could affect the accumulation of CtRLV or its symptomatology, in
152 particular when considering that it was systematically found associated with several umbraviruses
153 known to be synergistic with CtRLV and with other satellites or associated RNAs.
154

Compliance with ethical standards

Funding: This study was funded by the European Union through the Horizon 2020 Marie Skłodowska-Curie Actions Innovative Training Network (H2020 MSCA- 60 ITN) project "INEXTVIR" (GA 813542).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Data availability: The genome sequences for the hammerhead satellites reported here have been deposited in GenBank. The raw sequence datasets are available on request from the authors

Acknowledgments: The authors would like to thank the Platform GetPlaGe (INRAE, Toulouse, France) for the Illumina sequencing of the French samples.

References:

1. Roossinck MJ, Sleat D, Palukaitis P (1992) Satellite RNAs of plant viruses: structures and biological effects. Microbiol Rev 56:265-279. https://doi.org/mr.56.2.265-279.1992
2. Hu CC, Hsu YH, Lin NS (2009) Satellite RNAs and Satellite Viruses of Plants. Viruses 1:1325-1350. https://doi.org/10.3390/v1031325
3. Gnanasekaran P, Chakraborty S (2018) Biology of viral satellites and their role in pathogenesis. Curr Opin Virol 33:96-105. https://doi.org/10.1016/j.coviro.2018.08.002
4. Briddon RW, Ghabrial S, Lin NS, Palukaitis P, Scholthof KBG, Vetten, HJ (2012) Satellite and other virus-dependent nucleic acids. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus Taxonomy: classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Amsterdam, pp 1211-1219.
5. Branch AD, Robertson HD (1984) A replication cycle for viroids and other small infectious RNAs. Science 223:450-455. https://doi.org/10.1126/science.6197756
6. Bruening G, Passmore BK, van Tol H, Buzayan JM, Feldstein PA (1991) Replication of plant virus satellite RNA: evidence favors transcription of circular templates of both polarities. Mol Plant Microb Interact 4:219-225. https://doi.org/10.1094/mpmi-4-219
7. Bruening G (1989) Compilation of self-cleaving sequences from plant virus satellite RNAs and other sources. Methods Enzymol 180:546-558. https://doi.org/10.1016/0076-6879(89)80123-5
8. Navarro B, Rubino L, di Serio F (2017) Small Circular Satellite RNAs. In Viroids and Satellites, Hadidi A, Flores R, Randles JW, Palukaitis P, Eds, Academic Press, pp 659-669. https://doi.org/10.1016/B978-0-12-801498-1.00061-9
9. Miller WA, Hercus T, Waterhouse PM, Gerlach WL (1991). A satellite RNA of barley yellow dwarf virus contains a novel hammerhead structure in the self-cleavage domain. Virology, 183:711-720. https://doi.org/10.1016/0042-6822(91)91000-7
10. Domier LL (2012) Luteoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus Taxonomy: classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Amsterdam, pp 1045-1053.

11. Silver SL, Rasochova L, Dinesh-Kumar SP, Miller WA (1994) Replication of barley yellow dwarf virus satellite RNA transcripts in oat protoplasts. Virology 198:331-335. https://doi.org/10.1006/viro.1994.1036
12. Miller WA, Silver SL (1991) Alternative tertiary structure attenuates self-cleavage of the ribozyme in the satellite RNA of barley yellow dwarf virus. Nucl Acids Res 19:5313-5320. https://doi.org/10.1093/nar/19.19.5313
13. Song SI, Silver SL, Aulik MA, Rasochova L, Mohan BR, Miller WA (1999) Satellite cereal yellow dwarf virus-RPV (satRPV) RNA requires a double hammerhead for self-cleavage and an alternative structure for replication. J Mol Biol 293:781-793. https://doi.org/10.1006/jmbi.1999.3169
14. Song SI, Miller WA (2004) *Cis* and *trans* requirements for rolling circle replication of a satellite RNA. J VIROL 78:3072-3082. https://doi.org/10.1128/jvi.78.6.3072-3082.2004
15. Rasochová L, Miller WA (1996) Satellite RNA of barley yellow dwarf-RPV virus reduces accumulation of RPV helper virus RNA and attenuates RPV symptoms in oats. Mol Plant Microb Interact 9:646-650. https://doi.org/10.1094/MPMI-9-0646
16. Rasochová L, Passmore BK, Falk BW, Miller WA (1997) The satellite RNA of barley yellow dwarf virus-RPV is supported by beet western yellows virus in dicotyledonous protoplasts and plants. Virology 231:182-191. https://doi.org/10.1006/viro.1997.8532
17. Marais A, Faure C, Bergey B, Candresse T (2018) Viral double-stranded RNAs (dsRNAs) from plants: alternative nucleic acid substrates for high-throughput sequencing. Methods Molec Biol 1746:45-53. https://doi.org/10.1007/978-1-4939-7683-6_4

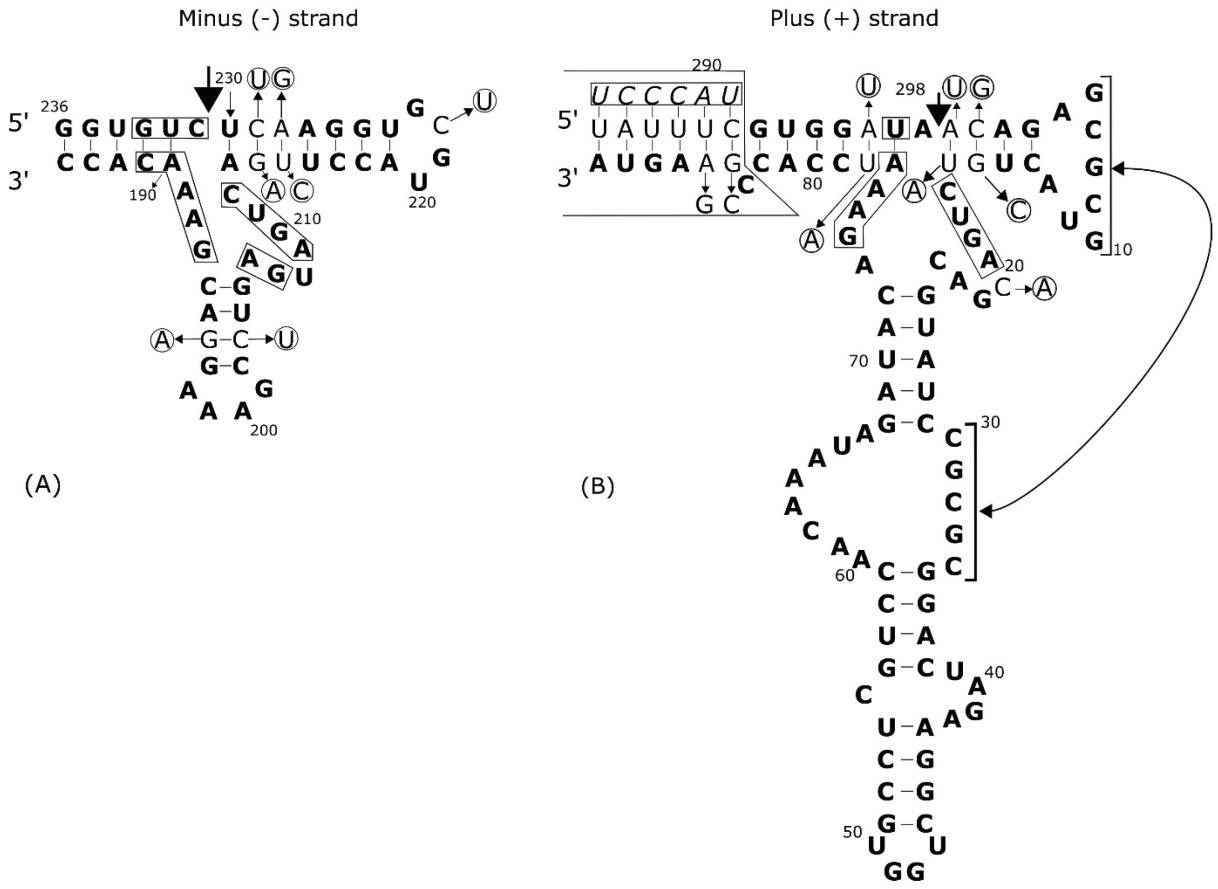

Figure 1. Proposed secondary structures for the minus and plus strands hammerhead ribozymes of carrot red leaf virus satellite 1 (CtRLV Sat1) based on the structure proposed for the ribozymes of CYDV-RPV satellite RNA. Nucleotide numbering is based on the full-length Sat1 RNA (298 nts). Boxed bases are conserved among satRNA hammerhead ribozymes. Big bold arrows indicate the cleavage site. Nucleotides in bold are conserved with CYDV-RPV satRNA, while small arrows and circled letters indicate mutations as compared to CYDV-RPV satRNA. (**A**) Minus strand ribozyme. Bases are numbered according to the complementary plus strand. (**B**) Plus strand ribozyme structure. Square brackets joined by a long double arrow indicate a pseudo-knot structure also found in CYDV-RPV satRNA.

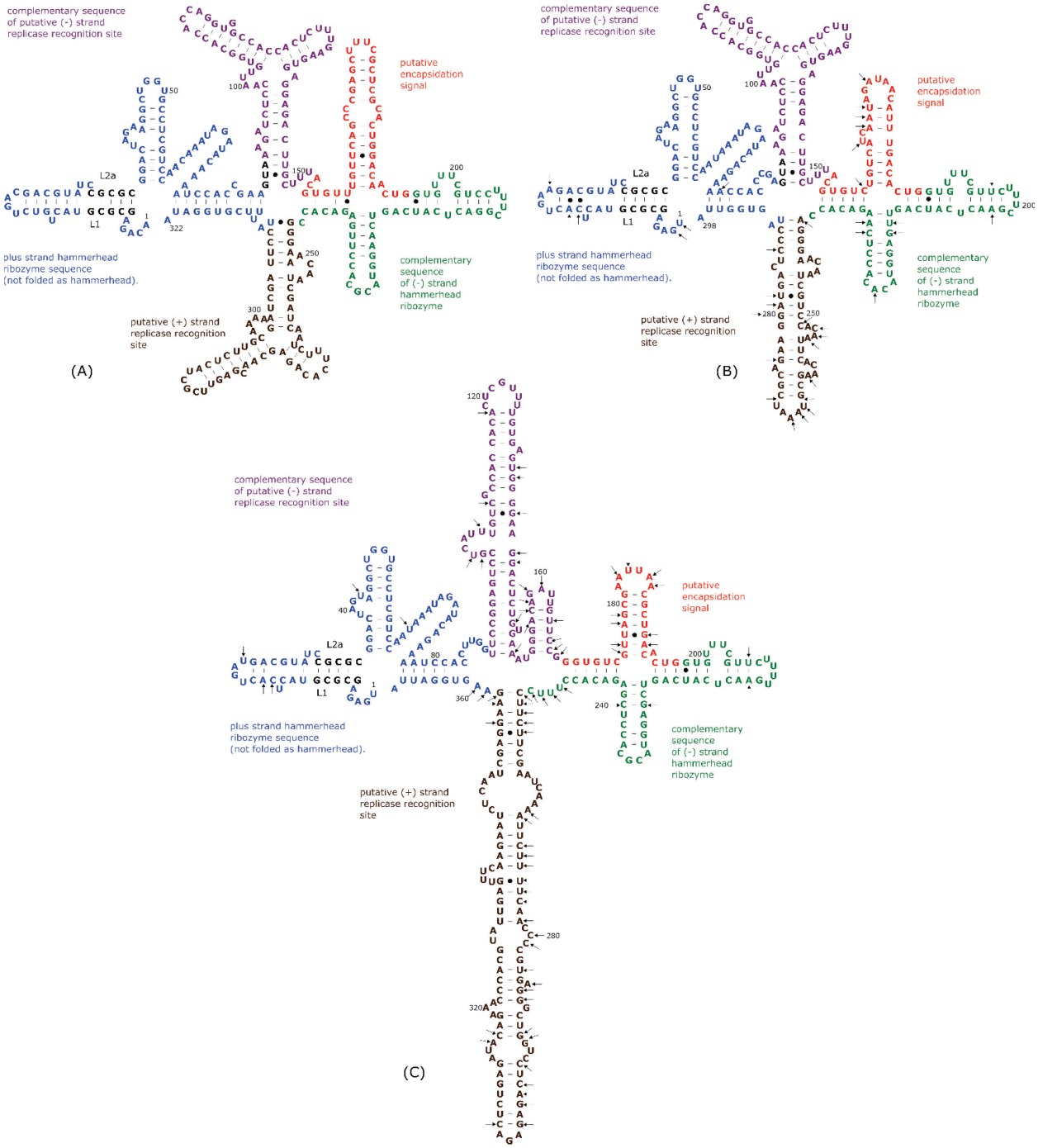

Figure 2. Known and putative functional domains in the most stable secondary structure of the plus strand of CYDV-RPV satRNA (**A**, adapted from Song & Miller, 2004), CtRLV Sat1 (**B**) and CtRLV Sat2 (**C**). Structural domains are color-coded for each known or proposed function as indicated, with the L1-L2a bases shown in black. Arrows indicate in Sat1 and Sat2 the nucleotide differences with respect to CYDV-RPV satRNA.