Molecular Plant Pathology

MOLECULAR PLANT PATHOLOGY (2011) **12**(7), 688–701 DOI: 10.1111/J.1364-3703.2010.00703.X

Mapping studies of the *Peach latent mosaic viroid* reveal novel structural features

AUDREY DUBÉ, FRANÇOIS BOLDUC, MARTIN BISAILLON AND JEAN-PIERRE PERREAULT*

RNA Group/Groupe ARN, Département de Biochimie, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, QC, Canada, J1H 5N4

SUMMARY

Knowledge of the structure of a viroid is critically important to elucidate the roles played by the various RNA motifs in the steps of the viroid's life cycle. A new technique, RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE), has recently been shown to be fast, reliable and applicable to the study of various RNA molecules. Consequently, this method was used to probe sequence variants of *Peach latent mosaic viroid* (PLMVd). Initially, probing data from RNA strands of both polarities of the Siberian C variant confirmed the secondary structures previously determined using both conventional and fastidious approaches. Subsequently, analysis of an Alberta variant showed an identical structure for the strand of (–) polarity, but the (+) polarity strand exhibited two differences from the Siberian C variant: the P11–L11 stem–loop domain formed a cruciform structure, and nucleotides from loops L1 and L11 were involved in the formation of a pseudoknot. The existence of both of these motifs was confirmed by site-directed mutagenesis. The subsequent probing of 12 natural sequence variants led to the elucidation of the criteria governing the formation of this novel pseudoknot. Importantly, this study revealed that the heterogeneity of a viroid is not limited to its nucleotide sequence, but may also occur at the structural level.

INTRODUCTION

Viroids are small (~250–400 nucleotides), single-stranded, circular RNAs that infect higher plants and cause significant losses in the agricultural industry (Flores *et al*., 2005). As viroids do not encode for any protein, the structures adopted by these infectious RNAs are essential for them to be able to accomplish the various activities of their life cycles. This is why, shortly after the discovery of viroids, research studies have largely focused on the determination of their structures (Riesner *et al*., 1979; Sanger *et al*., 1976; Steger *et al*., 1984). Clearly, the elucidation of the structures adopted by viroids is paramount in order to be able to understand the different mechanisms implicated in the viroid life cycle, processes such as transport, replication and pathogenesis, to name only a few examples. In general, the secondary structures of viroids have been predicted using computer software (Bussière *et al*., 1996). Clearly, the characterization of biological structures *in vitro* as well as *in vivo* is obviously more accurate for the elucidation of the structure–function relationship of a viroid. The secondary structure adopted by the *Potato spindle tuber viroid* (PSTVd) in solution was the first to be described based solely on biochemical approaches (Domdey *et al*., 1978; Gross *et al*., 1978). Specifically, the PSTVd circular monomeric strands were demonstrated to fold into a stable rod-like shape. Subsequently, the secondary structure of the *Peach latent mosaic viroid* (PLMVd) strands of (+) polarity in solution was determined using a combination of enzymatic mapping and oligonucleotide binding shift assays. This study revealed that a Siberian C sequence variant folded into a complex, branched secondary structure that included a pseudoknot (Bussière *et al*., 2000). This latter pseudoknot had never been predicted in any previous computer analysis. Recently, a similar approach was used to elucidate the secondary structure adopted by the counterpart PLMVd strands from the same sequence variant, that is to say those of (–) polarity (Dubé *et al*., 2010). This RNA molecule also adopted a branched secondary structure. However, several differences relative to the (+) strand were identified, including the absence of the pseudoknot found in the latter.

Because viroids are relatively long RNA species, the determination of their folded secondary structure remains fastidious. It is because of the inherent difficulty in elucidating their structure that most studies to date have only reported the secondary structure of a single variant (i.e. this is an achievement in itself). With the aim of determining the structure of several variants concurrently, and therefore taking advantage of the power of simultaneously analysing several sequences, a more efficient and rapid approach was needed. A global and rapid analysis of RNA structure is now possible as a result of the development of the

**Correspondence*: Email: jean-pierre.perreault@usherbrooke.ca

RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) method (Mortimer and Weeks, 2009). This method uses benzoyl cyanide (BzCN) which reacts with the 2′-hydroxyl groups of accessible single-stranded nucleotides. Base-paired or constrained nucleotides do not react with this hydroxyl-selective electrophile. During primer extension, the reaction is stopped at those locations that react, and the accessible nucleotide can then be identified. This technique is promising with the elucidation of the secondary structures of both a group I intron (Duncan and Weeks, 2008) and human immunodeficiency virus (HIV) RNAs (Wilkinson *et al*., 2008). In this report, the potential of the use of the SHAPE approach to elucidate the secondary structure of a viroid, used here as a model RNA species of several hundred nucleotides, is evaluated.

RESULTS AND DISCUSSION

SHAPE data confirm the structure of the PLMVd Siberian C variant

In order to evaluate the potential use of SHAPE for the structural characterization of viroids, this technique was applied to the Siberian C sequence variant whose structure has been determined previously using conventional methods (e.g. nuclease probing) (Bussière *et al*., 2000; Dubé *et al*., 2010). This PLMVd variant is also known as PLMVd number 24 from the *Prunus persica* Siberian C cultivar (accession number AF170496). According to the subviral database (Rocheleau and Pelchat, 2006), which includes the sequences of all viroids and related RNA species, it is known as PLMVd.034. In order to avoid any confusion, the name PLMVd.034 will be used for the balance of this article.

Nonradioactive, full-length PLMVd transcripts (338 nucleotides) of both (+) and (–) polarities were produced by *in vitro* transcription from the pPD1 plasmid that contains dimeric head-to-tail copies of PLMVd.034. The resulting concatemeric transcripts self-cleaved during synthesis, permitting the isolation of linear monomeric species after denaturing polyacrylamide gel electrophoresis (PAGE). After purification, the transcripts were dissolved in ultrapure water, heat denatured at 65 °C for 2 min and snap-cooled on ice for 5 min, before the addition of Tris-HCl/NaCl solution and incubation at 37 °C for 5 min. During the course of the present work, preliminary SHAPE probing was performed after slow and fast renaturation steps, and we found no significant difference. $MgCl_2$ and the reagent BzCN were then successively added to the solutions. As the reaction between the hydroxyl groups and BzCN occurs almost instantaneously (BzCN's half-life is 0.25 s at 37 °C in water), most of the remaining BzCN is inactivated by hydrolysis in a few seconds, forming a nonreactive product (Mortimer and Weeks, 2009). Consequently, the reactions do not need to be quenched. This fact also reduces the problem that is usually encountered with structural mapping approaches that require longer incubation times. After ethanol precipitation, the pellets were dissolved and primer extensions were performed with a radiolabelled primer. Under these conditions, the polymerase stops one nucleotide before that bearing the adduct formed with the BzCN reagent; hence, electrophoresis of the reverse transcription products on sequencing gels leads to the identification of the flexible residues. A typical gel for the analysis of reverse transcription reactions performed in the presence of an oligonucleotide complementary to the P4–L4 stem–loop region of the (+) polarity strand is shown in Fig. 1A.

The presence of metal ions, such as magnesium, can reduce the electrostatic repulsions between closely packed phosphate groups, thus stabilizing tertiary interactions (Misra and Draper, 2002; Soto *et al*., 2007). Therefore, the reactions were performed either in the absence or presence of 10 mM $MgCl_2$ in order to demonstrate the existence of certain tertiary interactions, such as the pseudoknot. The reactions performed without magnesium can highlight the absence of some interactions that normally require the presence of magnesium. In these cases, the residues involved in the formation of a pseudoknot should be more accessible in the absence of $MgCl_2$. After electrophoresis, the gels were exposed to phosphor screens which were then revealed using a Storm apparatus. The intensity of each band was subsequently analysed using SAFA software (Laederach *et al*., 2008) and represented graphically (Fig. 1C). In the present case, the addition of magnesium decreases the accessibility of some nucleotides of the L1 loop, suggesting that some rearrangements in this single-stranded domain occur on addition of magnesium.

As a typical primer extension reaction only permits the accurate analysis of a region spanning 50–100 nucleotides, and because a high level of resolution is required by SAFA software, four to five different oligonucleotides were required in order to probe the complete PLMVd sequence (Fig. 2A,B). Only the nucleotides located at the 3′-end that corresponded to the primer's binding site could not be probed. As the values of the SHAPE intensities can vary according to both the exposure time and half-life of the radioactive nucleotides used to label the primers, the intensities of the radioactivity between experiments were correlated with the values obtained for a common nucleotide in two consecutive runs performed with distinct primers. Data from at least triplicate reactions, in addition to the overlapping regions analysed using two distinct primers, were pooled together to establish the SHAPE-based secondary structures of the PLMVd RNA strands of both polarities [Fig. 2A,B; see Fig. S1 (Supporting Information) for the histograms of the SHAPE intensities of all positions for both polarities]. In each case, we have represented the structure according to previous experiments and assigned a colour to each nucleotide according to SHAPE data in order to show the differences between nuclease mapping and

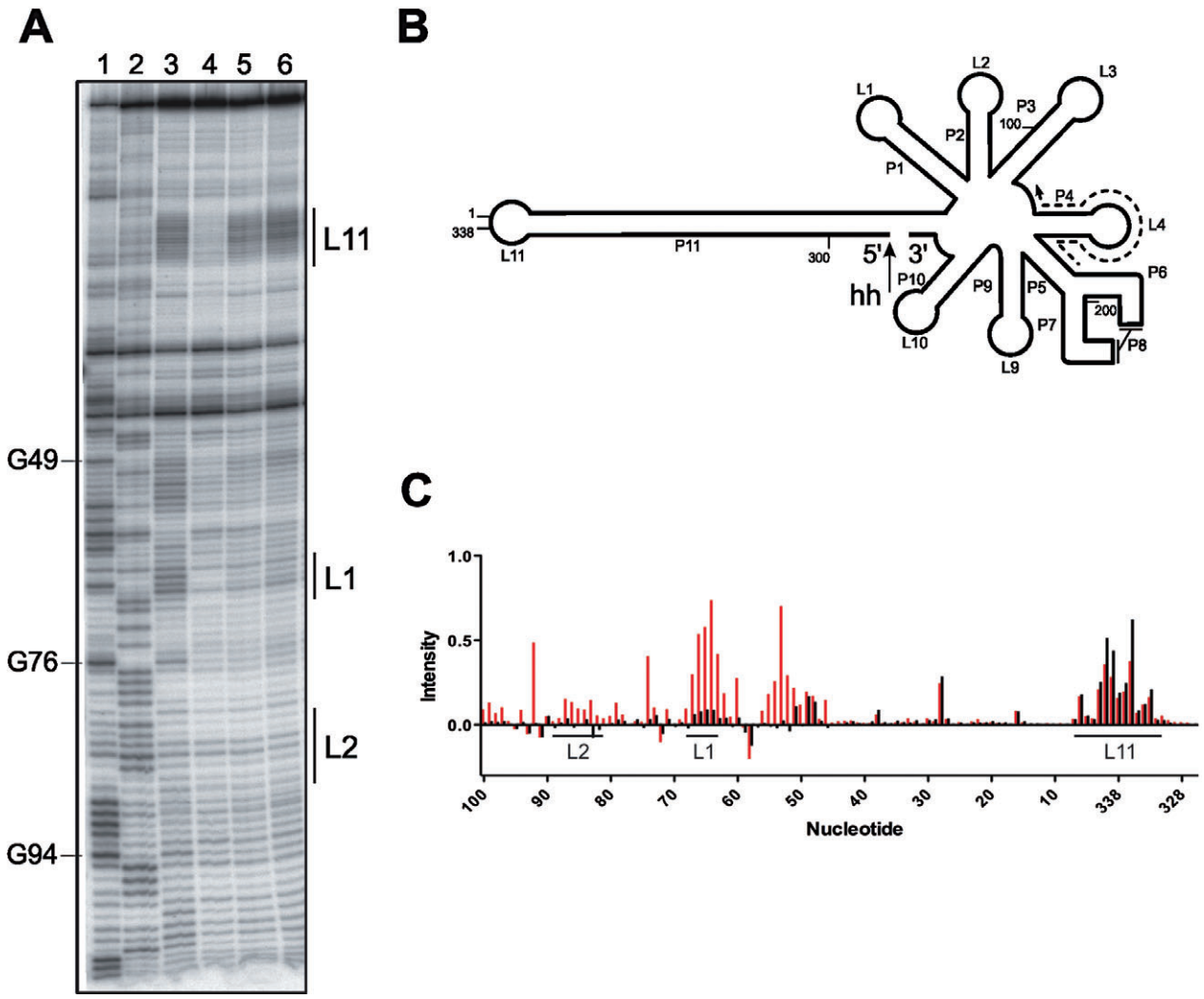

Fig. 1 Example of an RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) experiment probing the region from nucleotide 100 to nucleotide 328 of PLMVd.034. (A) A typical SHAPE experiment gel obtained with the full-length PLMVd.034(+) transcript. The primer extension was performed using a primer complementary to the P4–L4 stem–loop region. The first two lanes are primer extensions performed with dideoxy-cytosine and dideoxy-guanosine, respectively, in order to produce ladders. Lane 3 is a reaction performed in the absence of $MgCl_2$, but in the presence of benzoyl cyanide (BzCN). Lane 4 is a reaction performed in the presence of $MgCl_2$, but without BzCN. Finally, lanes 5 and 6 are duplicate reactions performed in the presence of both $MgCl_2$ and BzCN. Characteristic guanosine residues are indicated on the left of the gel, and the positions of three loops (L1, L2 and L11) are indicated on the right. (B) Schematic representation of the RNA template used for probing and the primer shown by broken lines. The arrow indicates the hammerhead self-cleavage site. (C) Graphical representation of the nucleotide accessibility as revealed by SAFA software analysis. The intensity of the control without BzCN was subtracted from the values for all nucleotides, and the results obtained for the complete *Peach latent mosaic viroid* (PLMVd) sequence were normalized using common nucleotides. The reactions performed with and without magnesium are represented by the black and red bars, respectively.

SHAPE probing. Most of the blue (less than 1% accessible) and yellow (less than 20% accessible) nucleotides are located in stems, whereas the orange (between 20% and 50% accessible) and red (more than 50% accessible) nucleotides are localized in loops or bulges. The results for the strands of both polarities are in relatively good agreement with the reported structures (Bussière *et al*., 2000; Dubé *et al*., 2010). However, some minor differences were observed for the bulged residues located along the P11 stem of both polarities that appeared to be inaccessible, most probably because they stacked with adjacent base-paired nucleotides. Similarly, in the (+) polarity strand, the residues of the L2 loop did not react, suggesting that they are inaccessible. In the latter case, there is additional information suggesting that this loop is probably buried inside the structure of the viroid, making it inaccessible to the probing reagent (A. Dubé and J. P. Perreault, unpublished data). Moreover, the sequence of the L2

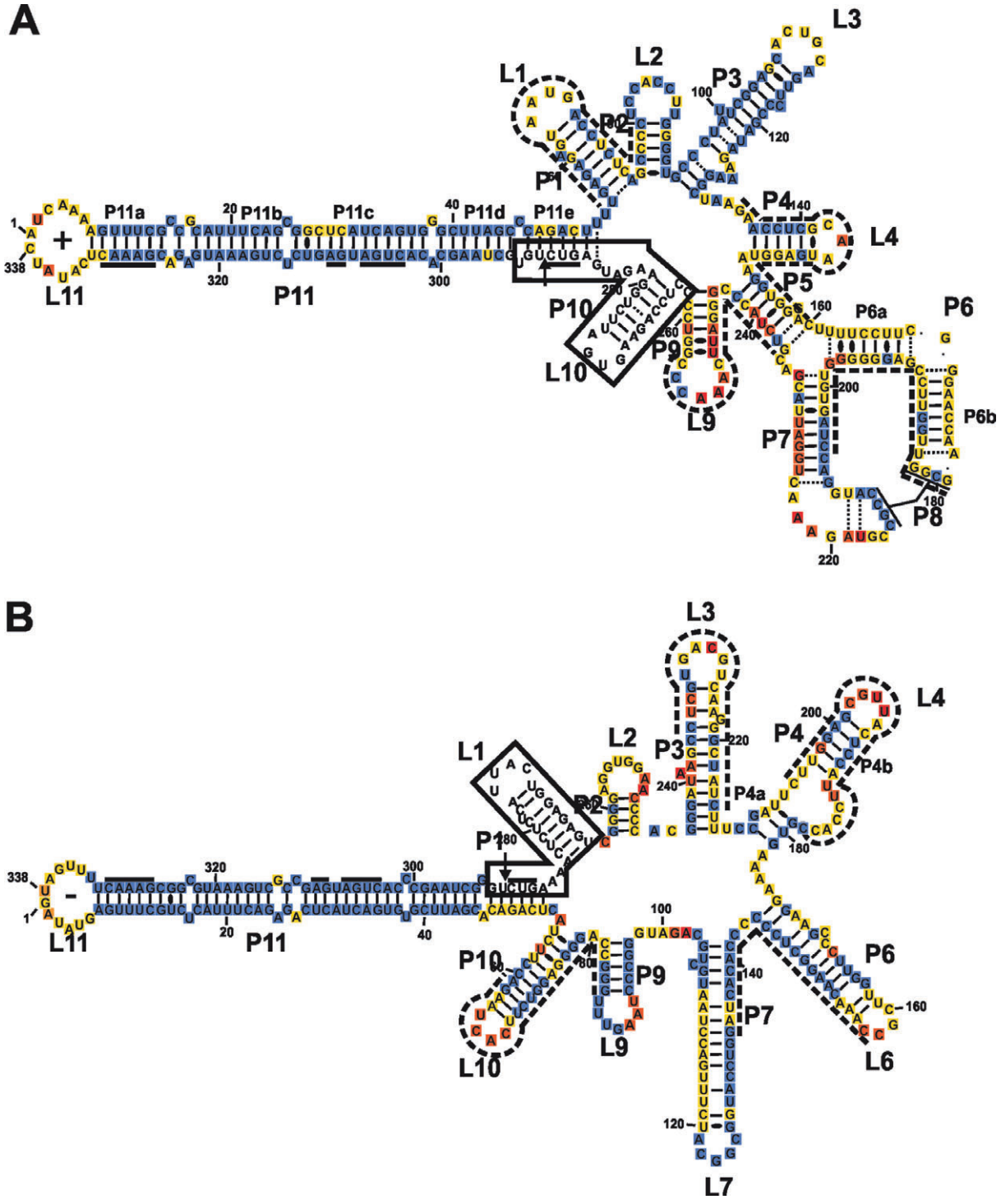

Fig. 2 RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) experiments elucidating the structure of the PLMVd.034 variant. (A) and (B) are the proposed secondary structures of the (+) and (–) polarities, according to previous nuclease mapping experiments. The flexibility of each nucleotide according to SHAPE probing from the present study is represented by colours. The blue boxes indicate nucleotides that are not accessible (i.e. intensities less than 1% of the highest value), and the yellow, orange and red boxes indicate levels of accessibility ranging from 1% to 20%, 20% to 50% and greater than 50% of the highest value, respectively. The boxed region at the 3′-end corresponds to the sequence complementary to the primer oligonucleotide used, and for which no result could be obtained. The sequences complementary to the other primers used are indicated by the broken lines. The arrows indicate the hammerhead self-cleavage site and the black boxes the conserved nucleotides that form the hammerhead motif. PLMVd, *Peach latent mosaic viroid*.

loop is perfectly conserved in all PLMVd sequence variants (>300 variants), suggesting that it is an important domain for an as yet unidentified biological activity. Conversely, there are also nucleotides from double-stranded regions, located in the region of the P5–P9 stems (from nucleotides 160 to 260), that show increased accessibility when compared with previous nuclease mapping, The differences between the two mapping methods can be explained by the fact that some regions could be buried inside the three-dimensional structure of PLMVd, rendering them inaccessible to nucleases, but accessible to BzCN which is a smaller molecule. In both cases, the analysis of the band pattern is not purely quantitative. When depicting the secondary structure based on RNase probing, the nucleotides at the base of a stem often appear as a mixture of intermediate states (including single- and double-stranded bases). Moreover, the first nucleotide in a loop (5′-side) appears to be partly double-stranded because it is stacked on the previous one, whereas it should be only single-stranded. In previous studies, we have found co-existing structures in PLMVd RNA using temperature gradient gel electrophoresis (Dubé *et al*., 2010). In addition, by software prediction tools and covariation analysis, we have shown that the region between stem P3 and P9 folds into various structures (Pelchat *et al*., 2000). The coexistence of more than one structure for this region (even in small proportions) may explain the subtle differences observed during the course of SHAPE probings. For both approaches, it is important to consider the sequence when performing the analysis, and many experiments are required to obtain a sufficient level of confidence. Thus, because different oligonucleotides for the primer extension analysis around the molecule have been used, the simple comparison of the intensity of the bands corresponding to nucleotides from different regions of the entire molecule is not straightforward. Moreover, minor differences can be indicative of local structural microheterogeneity. Finally, differences resulting from the presence of magnesium were also detected. For example, the P8 pseudoknot seems to require the presence of magnesium as, in the absence of $MgCl_2$, the loop L6 exhibited an important accessibility based on its reactivity with BzCN reagent (Fig. S1, top panel). More importantly, the SHAPE approach permits the validation of the PLMVd structures for the strands of both polarities through simple experiments that take only a few days, instead of the several months required by the classic techniques.

Analysis of the Alberta sequence variant

In order to further explore the potential of the SHAPE approach for the structural probing of viroids, another PLMVd sequence variant was analysed. Specifically, the Alberta variant, also known as variant 151.1 (accession number DQ680690), and referred to as PLMVd.282 (according to the subviral database), was probed. This sequence variant is 337 nucleotides in size, possesses 52 mutations when compared with PLMVd.034 and exhibits high infectivity when inoculated on the peach tree GF-305 cultivar (Dubé *et al*., 2010). Full-length transcripts of both polarities were produced *in vitro*, as described for the PLMVd.034 variant, and were then probed by SHAPE. The histograms summarizing the accessibility intensities of all positions for the strands of both polarities can be found in Fig. S2 (see Supporting Information). The secondary structures proposed, based on these data, for the RNA strands of both polarities are depicted in Fig. 3. For the PLMVd.282 variant of (+) polarity, both the P1–L1 and P11–L11 stem–loop structures, which compose the left-handed domain, showed significant differences from the PLMV.034 variant that are discussed below. Conversely, the right-handed domain, which is composed of the P2–L2 to P10–L10 stem–loop structures that include the P8 pseudoknot, was virtually identical to that of the PLMVd.034 variant, with the exception of minor structural rearrangements resulting from the sequence differences. Importantly, most of the PLMVd.282 RNA of (+) polarity folded into a secondary structure similar to that of the PLMVd.034 variant.

The left-handed domain includes four regions that exhibited different accessibility levels according to their reactivity with BzCN reagent. The nucleotides located in positions 18–23 and 314–319, which are largely complementary and form a double-stranded region according to the secondary structure of the PLMVd.034 variant of (+) polarity, appeared to be relatively accessible in the PLMVd.282 variant, whereas the adjacent sequences located on either side were not. This is a pattern indicative of the presence of an additional stem–loop on both the upper and lower strands of the P11 rod-like structure (Fig. 3A). Based on sequence comparisons, it has been proposed that the stem P11b and the adjacent nucleotides on both sides may adopt a slightly less stable, in terms of energy, alternative structure on both strands analogous to the hammerhead hairpin II (Ambros *et al*., 1998). This is in agreement with the accessibility pattern obtained by SHAPE. In the case of the PLMVd.034 variant, it was shown by nuclease mapping that a small fraction may have adopted this type of structure that includes the hairpins on both strands of the P11 stem (Bussière *et al*., 2000). According to SHAPE data, this second structure is favoured for the PLMVd.282 sequence variant. Therefore, the left-handed domain of the PLMVd.282 variant looks like a cruciform structure instead of having a long rod-like shape.

In contrast, several residues of both the L1 and L11 loops of the PLMVd.282 variant showed almost no reactivity to BzCN, indicating that they are no longer accessible and most probably are not in single-stranded regions, as is observed in the PLMVd.034 sequence (i.e. positions 335 to 2 and 64 to 68). More specifically, the sequences $_{335}$UAACC$_2$ and $_{64}$GGUUA$_{68}$ are complementary and may form a pseudoknot. The accessibility of these two regions was significantly greater in the absence of

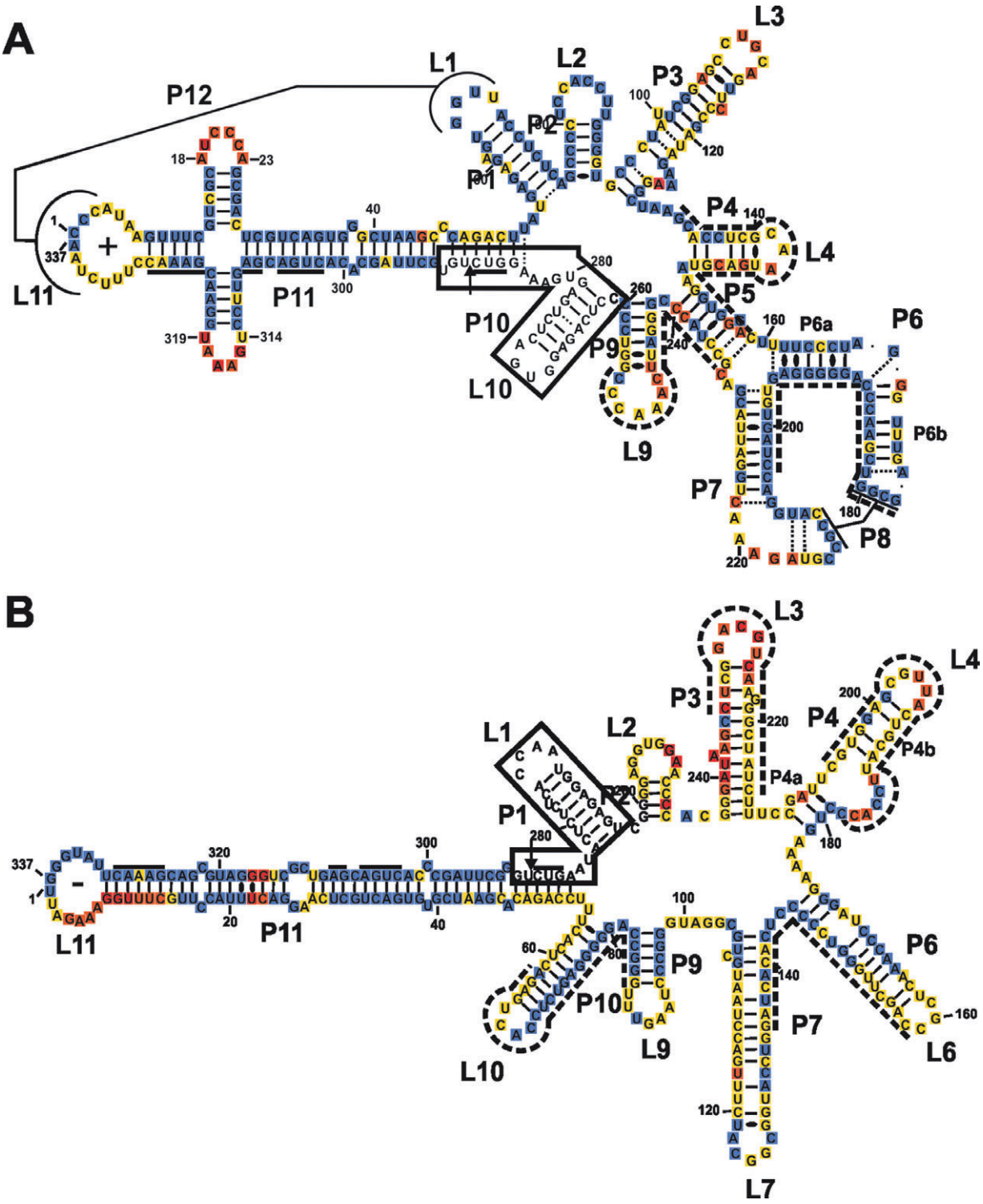

Fig. 3 Proposed secondary structures for both strands of the PLMVd.282 variant. The accessibility intensities are shown on the proposed secondary structures of both the (+) (A) and (–) (B) polarity strands of this variant. The boxed sequence shows the region covered by the first primer, for which no result could be obtained. The sequences complementary to the other primers used are indicated by the broken lines. The blue boxes indicate inaccessible nucleotides (i.e. intensities less than 1% of the highest value), and the yellow, orange and red boxes indicate different levels of accessibility (between 1% and 20%, between 20% and 50% and greater than 50% of the highest value, respectively). The arrows indicate the hammerhead self-cleavage site and the black boxes the conserved nucleotides that form the hammerhead motif. PLMVd, *Peach latent mosaic viroid*.

$MgCl_2$, suggesting the presence of a pseudoknot. This structure has been proposed previously based solely on sequence analysis (Ambros *et al*., 1998); however, this is the first time that it has received physical evidence based on probing in solution. This new stem is assigned the name of the P12 pseudoknot (previously it has been named P13; however, it is more appropriate to use P12 as there is no previous report of the presence of the previously named P12 structure). If P12 is indeed formed, it will generate an overall structure that includes two distinct pseudoknots (P8 and P12). Neither this additional stem–loop present in the rod-like domain, which provides a cruciform motif in the left-handed domain, nor the putative P12 pseudoknot, appeared to be present within the secondary structure adopted by the PLMVd strands of (–) polarity of the PLMVd.282 variant (Fig. 3B). Importantly, the use of the SHAPE approach has led to efficient probing of a new PLMVd sequence variant that reveals the presence of two new structural motifs. Moreover, the RNA strands of both the (+) and (–) polarities of the PLMVd.282 sequence variant fold differently, indicating that this phenomenon is not unique to the PLMVd.034 variant.

In the case of the PLMVd.282 strands of (–) polarity, SHAPE data indicated the presence of a secondary structure that was almost identical to that of the PLMVd.034 strand of the same polarity, although some minor differences were observed (e.g. the upper portion of the L11 loop). Neither the banding pattern supporting the formation of the cruciform, nor the P12 pseudoknot, was detected (Fig. S2). Thus, PLMVd RNA strands of (–) polarity appeared to fold into similar secondary structures even though they differed in terms of nucleotide sequences (52 mutations). It is also interesting to note that the sequence variations influenced the folding of the (+) polarity strands, but seemed to have only a limited and local effect on the secondary structures of the (–) polarity strands.

Confirmation of the cruciform motif

The cruciform structure detected for the PLMVd.282 RNA strands of (+) polarity corresponds to the formation of the hairpin II of the hammerhead structure on the lower strand, and to its counterpart on the upper strand (Fig. 4A,B). Both stems are located in the middle of the P11 rod-like structure. Covariation of the base-paired nucleotides forming this hammerhead helix II has been reported in all PLMVd sequencing studies (Ambros *et al*., 1998; Fekih Hassen *et al*., 2007; Pelchat *et al*., 2000). This supports its existence in all PLMVd sequence variants and is in agreement with the existence of a selective pressure in favour of the hammerhead self-cleavage activity. Bioinformatic predictions of the secondary structures of variants PLMVd.034 and PLMVd.282 were performed in order to determine from the probabilities whether the inclusion of the hairpin II that produces a cruciform structure may be detected instead of a long rod-like structure for the P11–L11 stem–loop domain. When several secondary structure predictions were analysed, some including the cruciform were obtained for PLMVd.282, although those with P11 forming a rod-like structure were favoured in terms of energy (data not shown). Conversely, none including the cruciform were retrieved for the PLMVd.034 variant. Refinement of the analysis using only the sequence of the P11–L11 stem–loop domain did not help to elucidate the rules governing the formation of the long rod-like versus cruciform motifs, because the differences in terms of energy were not sufficiently significant. Prediction of the secondary structures of many other PLMVd sequence variants revealed that only a small proportion should favour inclusion of the hammerhead hairpin II within the most stable secondary structures.

Subsequently, a minimal mutational study was performed in order to determine whether or not the hairpin II was included within the rod-like structure formed by the P11–L11 stem–loop. Initially, full-length PLMVd.282 was trimmed to 174 nucleotides in order to isolate the left-handed domain (i.e. P10, P11, P1 and P2 stems; see Fig. 4C). For every molecule of (+) polarity, the nucleotide U_9 was changed to A and A_{330} to U in order to prevent cleavage by the hammerhead ribozyme. The complementary inversion was performed on the (–) polarity molecule. Covariable mutations were also added to both P2 and P9 of both the (+) and (–) polarities in order to lower the stabilities of these stems and thereby facilitate the annealing of the primer used during the reverse transcription step. When the corresponding transcripts were synthesized and analysed by the SHAPE approach, the probing data were virtually identical to those obtained from the full-length sequence (Figs 4C and S2A). Interestingly, it is easier to analyse the difference in terms of accessibility when probing a short transcript in comparison with the full-length molecule. Furthermore, another difficulty of probing longer transcripts is the fact that it requires the use of various oligonucleotides for the primer extension which complicates the analysis, whereas the experiment probing the small transcript requires a unique primer extension reaction to cover the entire molecule. Briefly, the residues involved in the formation of the hammerhead hairpin II on the lower strand, and the corresponding residues of the upper strand, were either inaccessible or showed only a limited accessibility, whereas those of the two loops exhibited a significant reactivity with BzCN. The residues involved in the formation of the P12 pseudoknot also showed no, or only limited, reactivity with the SHAPE reagent, supporting the formation of this helix. Importantly, these data showed that the small transcripts folded exactly according to the structure probed for the full-length transcript, and therefore could be used for further studies.

Next, a sequence mutant that should favour the formation of a long rod-like P11 stem was produced. In order to accomplish this, the residues located in positions 13–17 and 320–324 (i.e. $_{13}$GUCGC$_{17}$ and $_{320}$GGAAC$_{324}$), which are involved in the

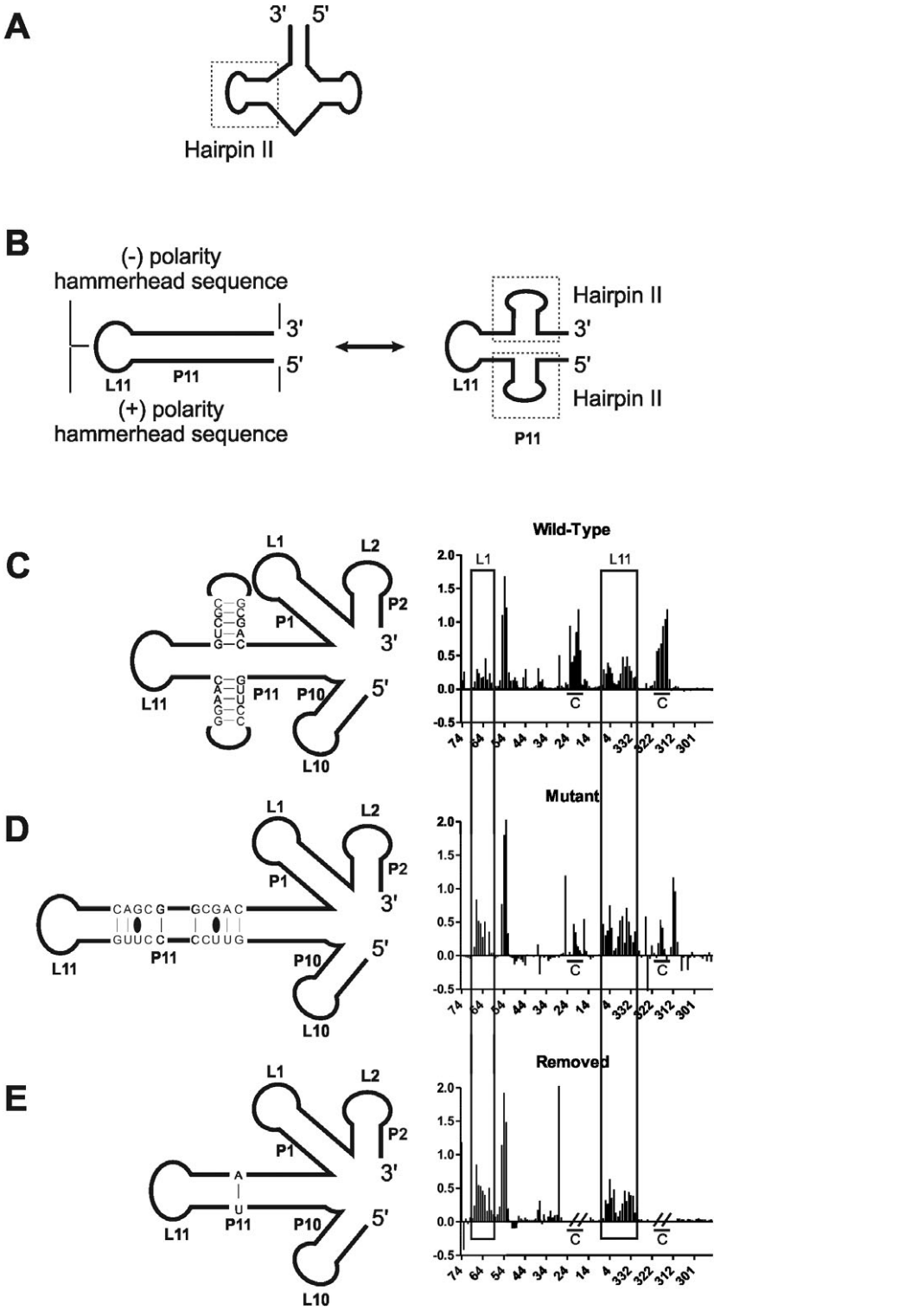

Fig. 4 Analysis of the cruciform motif. General structure of a hammerhead ribozyme (A). The P11 stem can fold into either the P11 stem or into the cruciform motif which includes the hairpin II of the hammerhead ribozymes of both strands (B). RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) profile for the wild-type sequence (C) compared with that of a mutated sequence that prevents the formation of the cruciform motif, but keeps the P11 stem–loop (D). In (E), the sequences implicated in the formation of the cruciform were replaced by one base pair. For each graph, the regions corresponding to nucleotides of L1 and L11 are boxed, and those corresponding to the nucleotides of the cruciform loop structure are underlined and identified by the letter 'c'.

formation of the hairpin II and in the counterpart helix, were mutated in such a way as to favour the formation of the long rod-like P11 stem (i.e. $_{13}$CAGCG$_{17}$ and $_{320}$CCUUG$_{324}$, respectively). Our strategy involves the mutation of all positions, rather than only some, in order to ensure that there is no potential ambiguity in the analysis. Probing of the corresponding transcript showed that the residues of the two loops in question were less accessible for the BzCN reaction, supporting the idea that this transcript folded into a long, rod-like structure instead of into the cruciform motif. Indeed, the probing data were in agreement with those obtained for the equivalent region of the PLMVd.034 sequence variant. Interestingly, the nucleotides suspected of being involved in the formation of the P12 pseudoknot did not react with the BzCN reagent (Fig. 4D). This is a good indication that the formation of the P12 pseudoknot is at least partially independent of the presence of the cruciform motif. Finally, a mutant in which the sequences involved in the formation of the hairpin II on the lower strand, and those involved in its counterpart stem–loop motif on the upper strand, were deleted was also synthesized (Fig. 4E). More specifically, these nucleotides were replaced by a single base pair in order to compensate for the spacing created by the presence of the palindromic stem–loop structures. The probing data for this mutant indicated that it folded into a rod-like structure (avoiding the inclusion of the hairpin II and its counterpart) and that it included the P12 pseudoknot. Thus, the formation of the cruciform is not essential for the formation of the P12 pseudoknot, and appears to be less stable than P11 according to the bioinformatic predictions.

Characterization of the P12 pseudoknot

Both the full-length and smaller PLMVd.282-derived transcripts of (+) polarity folded into structures that permitted the detection of SHAPE banding patterns that were in agreement with the formation of the P12 pseudoknot (Figs 3A and 4C). The formation of this pseudoknot involved base-pairing between the nucleotides located in positions 335 to 2 of the L11 loop with those located in positions 64–68 of the P1 loop, in the case of the PLMVd.282 variant (Fig. 5A). In order to obtain additional physical support for the presence of the P12 pseudoknot, a mutation analysis was performed using the smaller versions of 174 nucleotides. Initially, a mutant that did not permit the formation of the P12 pseudoknot was produced by changing the sequence of $_{335}$UAACC$_2$ to $_{335}$AUUGG$_2$ (Fig. 5B). The resulting SHAPE banding pattern showed that the putative nucleotides involved in the P12 pseudoknot exhibited a drastically improved reactivity with the BzCN reagent in the presence of $MgCl_2$. Clearly, these residues were located in a single-stranded region in this construction. When the compensating mutations were introduced in the L1 loop, the SHAPE banding pattern observed was in agreement with the formation of the P12 pseudoknot (Fig. 5C). However, the 'rescue' exhibited a BzCN reactivity that was not as low as that observed for the wild-type sequences, indicating that the P12 pseudoknot could be formed by more than one nucleotide sequence. Consequently, the stability of the P12 pseudoknot can change and is determined by the nucleotide sequence involved.

In order to further identify the nucleotides that may be involved in the formation of the P12 pseudoknot, a covariation analysis was performed using small RNA molecules of 174 nucleotides. All available PLMVd sequences of (+) polarity that did not contain the insertion specific for peach calico disease (Malfitano *et al*., 2003) were aligned. The sequences of both the L1 and L11 regions were then isolated and the potential of them hybridizing together was analysed using RNAhybrid software (Rehmsmeier *et al*., 2004). A total of 306 variants was analysed and classified according to the minimal free energy (MFE) of hybridization of the most stable interaction between nucleotides of L1 and L11 (Fig. S3, see Supporting Information). The MFE of hybridization of each pair of sequences given by RNAhybrid ranged from −13.2 to −2.5 kcal/mol.

Subsequently, in order to verify whether or not this putative pseudoknot was indeed formed in solution, and if it was possible to elucidate rules governing its formation, transcripts similar to the small PLMVd.282, including the sequence found in nature, were synthesized and probed by SHAPE. Specifically, 12 different sequences were chosen based on their MFE of hybridization. These variants represented different situations (e.g. low or high values of MFE). All transcripts were probed by SHAPE either in the presence or absence of magnesium ions; the data are summarized in Table 1. Figure 6 presents the proposed secondary structures, including both the L1 and L11 loops that are not base-paired together. Briefly, the formation of the P12 pseudoknot was detected with the transcripts derived from variants possessing an MFE of hybridization of less than −9.0 kcal/mol. Conversely, no P12 pseudoknot could be detected for the transcripts possessing an MFE value of more than −8.0 kcal/mol. The transcripts for which a median MFE value was calculated from RNAhybrid (those ranging from −9.0 to −8.0 kcal/mol) constitute a more ambiguous situation. The SHAPE banding pattern suggested the presence of the P12 pseudoknot within the sequence derived from the PLMVd.011 variant (MFE = −8.3 kcal/mol), but not in the case of the PLMVd.254 variant (MFE = −8.6 kcal/mol). The only exception observed was the PLMVd.093 variant for which an MFE value of −12.7 kcal/mol was calculated, but the SHAPE banding pattern did not support the formation of the pseudoknot. However, it is reasonable to suggest that the localization of the interacting nucleotides on L11 in this variant is detrimental to P12 formation. More specifically, most of the nucleotides that could be involved in base-pairing were located on the lower face of the L11 loop, which most probably is not a favourable situation for the formation of the pseudoknot (Fig. 6). All of the other variants had SHAPE

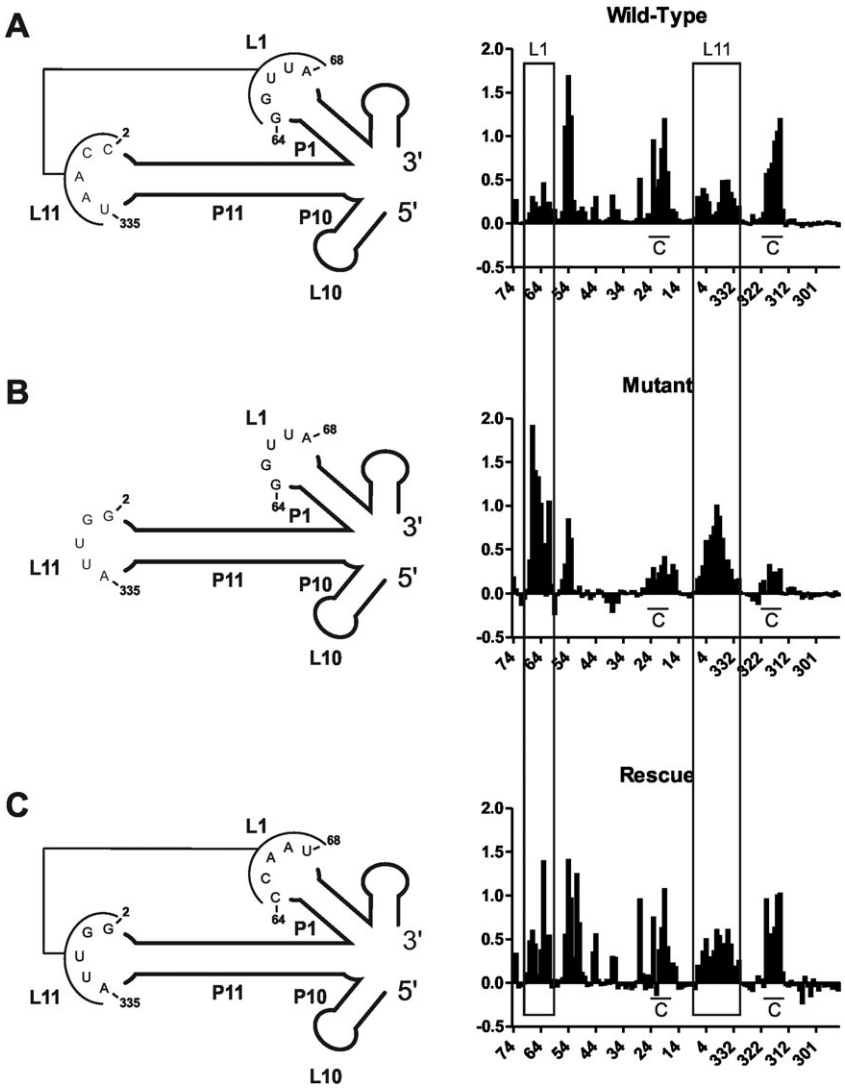

Fig. 5 Characterization of the P12 pseudoknot using a small model transcript derived from PLMVd.282. RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) analysis of the wild-type (A), mutant (B) and rescue (C) small transcripts derived from PLMVd.282. The nucleotides implicated in the interaction are indicated on the secondary structures. The intensities of the primer extension reactions for each nucleotide are represented graphically. The L1 and L11 loops are boxed, and the nucleotides implicated in the cruciform loop are underlined and indicated with the letter 'c'. PLMVd, *Peach latent mosaic viroid*.

Table 1 Analysis of *Peach latent mosaic viroid* (PLMVd) variants for the presence of the P12 pseudoknot and the cruciform motif.

			SHAPE probing	
Variant	RNA hybrid minimal free energy (kcal/mol)	Location on L11	Presence of P12	Presence of cruciform
PLMVd.261	−12.8	Near	Yes	Yes
PLMVd.093	−12.7	Far	L1	Yes
PLMVd.042	−9.8	Middle	Yes	Yes
PLMVd.282	−9.7	Middle	Yes	Yes
PLMVd.232	−9.4	Middle	Yes	Yes
PLMVd.015	−9.1	Near	Yes	Yes
PLMVd.254	−8.6	Near	No	Yes
PLMVd.011	−8.3	Middle	Yes	Yes
PLMVd.013	−7.1	Middle	No	Yes
PLMVd.034	−7.1	Middle	L1	No
PLMVd.009	−6.9	Far	No	Yes
PLMVd.276	−3.6	—	No	No

SHAPE, RNA-selective 2′-hydroxyl acylation analysed by primer extension.

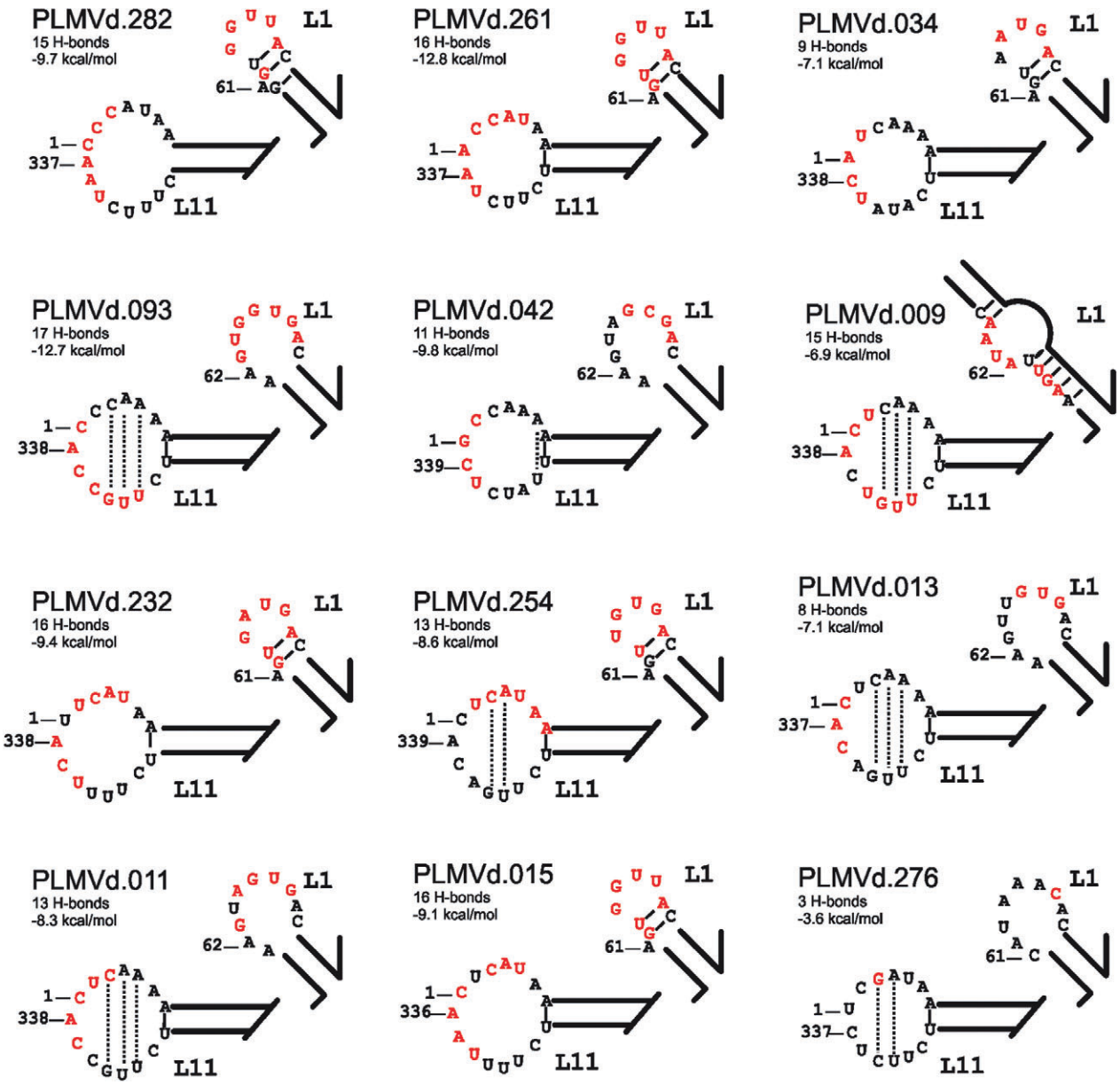

Fig. 6 Structure of both the L1 and L11 loops of natural small *Peach latent mosaic viroid* (PLMVd) variants. The sequence of each full-length variant was subjected to mFold analysis (Zuker, 2003), and only the structures of L1 and L11 are represented. The nucleotides predicted to be involved in the formation of the P12 pseudoknot are highlighted in red. The numbering of each molecule is according to that of the full-length natural variant. Possible base-pair interactions within the loop L11 are indicated by the dotted lines.

banding patterns that were in agreement with the formation of the P12 pseudoknot, involving at least a few nucleotides located on either the middle or upper faces of the L11 loop, most probably so as to favour the occurrence of the interaction. In the case of the two transcripts that showed SHAPE banding patterns that did not support the presence of the P12 pseudoknot, an increase in accessibility in the region of L1 in the absence of magnesium ions, reminiscent of a possible tertiary interaction, remains to be observed for the nucleotides located in the L1 loop. This was the case for the transcripts derived from both the PLMVd.034 and PLMVd.093 variants. However, the analysis of the complete SHAPE banding pattern did not reveal the existence of any possible region with which these nucleotides could interact. Otherwise, analysis of the sequences of (+) polarity of all available PLMVd variants revealed that 215 variants (70.3%) had an MFE below –9.0 kcal/mol, suggesting that they favour the formation of the P12 pseudoknot (Fig. S3). Conversely, 77 variants (25.1%) should not contain this pseudoknot as they possess an

MFE above –8.0 kcal/mol, and only 14 variants (4.6%) had MFE in the region of uncertainty (i.e. –9.0 to –8.0 kcal/mol).

Finally, the SHAPE banding pattern confirmed the formation of the cruciform in all variants probed, except for variants PLMVd.034 and PLMVd.276. Interestingly, these two variants cannot form the P12 pseudoknot. At this stage, it would be tempting to hypothesize that the cruciform structure is essential for the formation of the pseudoknot. However, as demonstrated previously (Fig. 4), the absence of the cruciform does not necessarily prevent the formation of the P12 pseudoknot. Finally, the presence of the cruciform does not imply the formation of the P12 pseudoknot, as variants PLMVd.009, PLMVd.013, PLMVd.093 and PLMVd.254 possess a cruciform, but not the P12 pseudoknot.

In this article, the SHAPE approach was used successfully to confirm the folded secondary structure of both strands of the PLMVd variant PLMVd.034, as well as to determine that of the variant PLMVd.282. The analysis performed led to the characterization of a novel pseudoknot, as well as the documentation of the formation of both long rod-like and cruciform motifs by the sequences of the P11–L11 domains of the various sequence variants. It would be of great interest to clarify the rules governing the formation of these different motifs, as well as to determine their biological importance. Essentially, this study revealed that the heterogeneity of a viroid is not limited to the sequence level, but instead can also occur at the structural level. Moreover, SHAPE appears to be a rapid, quantitative and reliable approach for the determination of the secondary structures of viroids and, more generally, of long RNA molecules. Lastly, with respect to PLMVd, it is now clear that the strands of both polarities fold into different secondary structures, and that this diversity is also observed within the (+) polarity genomes of the entire population of PLMVd.

EXPERIMENTAL PROCEDURES

In vitro transcription and purification of RNAs

Transcription reactions for both polarities of PLMVd variant PLMVd.034 were performed as described previously from the pPD1 plasmid (Dubé *et al*., 2010). The variant PLMVd.282 was dimerized and then cloned into pGEM-T (Promega, Madison, WI, USA). Plasmids containing dimers inserted in both orientations were conserved, and transcription reactions were performed on *Spe*I-digested plasmids with purified T7 RNA polymerase. All transcription reactions were fractionated by denaturing (8 M urea) 5% PAGE (19 : 1 ratio of acrylamide/bisacrylamide) using 1 × TBE (89 mM Tris, 89 mM boric acid, pH 8.0) as running buffer. The self-cleaved 338-nucleotide RNA products were excised under UV shadowing and eluted overnight in elution buffer [500 mM NH_4OAc, 1 mM ethylenediaminetetraacetic acid (EDTA) and 0.1% sodium dodecylsulphate (SDS)]. The RNA products were then ethanol precipitated, washed with ethanol 70% and quantified by spectrophotometry.

For small PLMVd molecules, long DNA oligonucleotides containing the inverse sequence of the T7 promoter added to the 5′ end of the DNA sequence of the PLMVd molecule in question were used as templates. DNA templates were prepared by polymerase chain reaction (PCR) filling reactions using *Pwo* DNA polymerase (Roche Diagnostics, Indianapolis, IN, USA) and an oligonucleotide that included the T7 RNA promoter's complementary sequence. The oligonucleotides used are presented in Fig. S4 (see Supporting Information). The transcriptions were then performed as described above.

5′-end labelling of oligonucleotides

Oligonucleotides (10 pmol) complementary to the stem–loops P1, P3, P4, P7, P9 and P10 of strands of both polarities of variants PLMVd.034 and PLMVd.282 were 5′-end labelled in the presence of 3.2 pmol [γ-^{32}P]-ATP (6000 Ci/mmol, New England Nuclear, Waltham, MA, USA) and 3 U of T4 polynucleotide kinase according to the manufacturer's recommended procedure (USB, Santa Clara, CA, USA). The reactions were performed at 37 °C for 60 min. The labelled oligonucleotides were fractionated by denaturing (8 M urea) 10% PAGE (19 : 1 ratio of acrylamide/bisacrylamide) using 1 × TBE as running buffer. After autoradiography, the bands containing the appropriate 5′-end-labelled oligonucleotides were excised and DNA was recovered as described above. For the analysis of the small PLMVd sequences, oligonucleotides complementary to P2 for the (+) strand and P9 for the (–) strand were used. The same P2 oligonucleotide was used for all variants, with the exception of variants PLMVd.261, PLMVd.232, PLMVd.093 and PLMVd.009, which have some differences in this region. All of the oligonucleotides used are shown in Fig. S5 (see Supporting Information).

SHAPE

SHAPE was performed on RNA aliquots of 5 pmol (diluted in water in order to adjust the volume to 8 μL). The samples were heated at 65 °C for 2 min, put on ice for 5 min and then 1 μL of folding buffer was added (500 mM Tris-HCl, pH 7.5, 500 mM NaCl). The mixtures were placed at 37 °C for 5 min and then 1 μL of 100 mM $MgCl_2$ was added and incubated at 37 °C for 10 min. Finally, 1 μL of 600 mM BzCN in dimethylsulfoxide (DMSO) was added. There was no need to stop the reaction as the BzCN reaction is instantaneous and any remaining reagent will be hydrolysed. All reactions were then ethanol precipitated in the presence of glycogen and washed with ethanol 70%. Negative controls with DMSO, but without BzCN and magnesium, were also performed under identical conditions.

Primer extensions were performed using SuperScript III (Invitrogen, Carlsbad, CA, USA) and radiolabelled oligonucleotide

complementary to the RNA in question. First, all pellets from the above reactions were dissolved in 12 μL of water and then 1 μL (1 pmol) of labelled oligonucleotide was added. The resulting mixtures were then incubated for 5 min at 65 °C, followed by 5 min at 37 °C and 1 min at 4 °C. A mixture of 4 μL of 5 × first strand buffer, 1 μL dithiothreitol (DTT) (0.1 M), 1 μL deoxynucleoside triphosphates (dNTPs) (10 mM) and 2 μL DMSO was added to each sample and the reactions were incubated at 61 °C for 1 min. Controls with either 1 μL of 10 mM dideoxy-cytidine triphosphate (ddCTP) or 2 μL of 10 mM dideoxy-guanosine triphosphate (ddGTP) were also performed. Finally, 0.5 μL of Superscript III enzyme was added and extension reactions were incubated at 61 °C for 10 min prior to the addition of 2 μL of NaOH (2 M). The reactions were then incubated at 95 °C for 5 min and ethanol precipitated with glycogen. After washing with 70% ethanol, the samples were dissolved in 10 μL of 95% formamide, 10 mM EDTA and 0.025% xylene cyanol solution and the radioactivity was evaluated by Cerenkov counting. Equivalent amounts of radioactivity were electrophoresed on denaturing 8% PAGE gels, which were subsequently dried and visualized by exposure to phosphor imaging screens. Every gel was analysed using SAFA software (http://safa.stanford.edu) (Laederach *et al*., 2008). Each experiment was repeated three times, and the results are presented as an average of the three experiments. Finally, the SHAPE reactivities were normalized to a uniform scale using a simple normalization scheme described previously (Low and Weeks, 2010).

ACKNOWLEDGEMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council (NSERC Canada, grant number 155219-07) to J.P.P. The RNA group is supported by grants from the Université de Sherbrooke and the Canadian Institutes of Health Research (CIHR, grant number PRG-80169). A.D. was the recipient of predoctoral fellowships from both the NSERC and Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT). J.P.P. holds the Canada Research Chair in Genomics and Catalytic RNA. M.B and J.P.P. are members of the Centre de Recherche Clinique Étienne-Lebel.

REFERENCES

Ambros, S., Hernandez, C., Desvignes, J.C. and Flores, R. (1998) Genomic structure of three phenotypically different isolates of peach latent mosaic viroid: implications of the existence of constraints limiting the heterogeneity of viroid quasispecies. *J. Virol.* **72**, 7397–7406.

Bussière, F., Lafontaine, D. and Perreault, J.P. (1996) Compilation and analysis of viroid and viroid-like RNA sequences. *Nucleic Acids Res.* **24**, 1793–1798.

Bussière, F., Ouellet, J., Cote, F., Levesque, D. and Perreault, J.P. (2000) Mapping in solution shows the peach latent mosaic viroid to possess a new pseudoknot in a complex, branched secondary structure. *J. Virol.* **74**, 2647–2654.

Domdey, H., Jank, P., Sanger, L. and Gross, H.J. (1978) Studies on the primary and secondary structure of potato spindle tuber viroid: products of digestion with ribonuclease A and ribonuclease T1, and modification with bisulfite. *Nucleic Acids Res.* **5**, 1221–1236.

Dubé, A., Baumstark, T., Bisaillon, M. and Perreault, J.P. (2010) The RNA strands of the plus and minus polarities of peach latent mosaic viroid fold into different structures. *RNA*, **16**, 463–473.

Duncan, C.D. and Weeks, K.M. (2008) SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry*, **47**, 8504–8513.

Fekih Hassen, I., Massart, S., Motard, J., Roussel, S., Parisi, O., Kummert, J., Fakhfakh, H., Marrakchi, M., Perreault, J.P. and Jijakli, M.H. (2007) Molecular features of new Peach Latent Mosaic Viroid variants suggest that recombination may have contributed to the evolution of this infectious RNA. *Virology*, **360**, 50–57.

Flores, R., Hernandez, C., Martinez de Alba, A.E., Daros, J.A. and Di Serio, F. (2005) Viroids and viroid–host interactions. *Annu. Rev. Phytopathol.* **43**, 117–139.

Gross, H.J., Domdey, H., Lossow, C., Jank, P., Raba, M., Alberty, H. and Sanger, H.L. (1978) Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature*, **273**, 203–208.

Laederach, A., Das, R., Vicens, Q., Pearlman, S.M., Brenowitz, M., Herschlag, D. and Altman, R.B. (2008) Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nat. Protoc.* **3**, 1395–1401.

Low, J.T. and Weeks, K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150–158.

Malfitano, M., Di Serio, F., Covelli, L., Ragozzino, A., Hernandez, C. and Flores, R. (2003) Peach latent mosaic viroid variants inducing peach calico (extreme chlorosis) contain a characteristic insertion that is responsible for this symptomatology. *Virology*, **313**, 492–501.

Misra, V.K. and Draper, D.E. (2002) The linkage between magnesium binding and RNA folding. *J. Mol. Biol.* **317**, 507–521.

Mortimer, S.A. and Weeks, K.M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* **4**, 1413–1421.

Pelchat, M., Levesque, D., Ouellet, J., Laurendeau, S., Levesque, S., Lehoux, J., Thompson, D.A., Eastwell, K.C., Skrzeczkowski, L.J. and Perreault, J.P. (2000) Sequencing of peach latent mosaic viroid variants from nine North American peach cultivars shows that this RNA folds into a complex secondary structure. *Virology*, **271**, 37–45.

Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.

Riesner, D., Henco, K., Rokohl, U., Klotz, G., Kleinschmidt, A.K., Domdey, H., Jank, P., Gross, H.J. and Sanger, H.L. (1979) Structure and structure formation of viroids. *J. Mol. Biol.* **133**, 85–115.

Rocheleau, L. and Pelchat, M. (2006) The Subviral RNA Database: a toolbox for viroids, the hepatitis delta virus and satellite RNAs research. *BMC Microbiol.* **6**, 24.

Sanger, H.L., Klotz, G., Riesner, D., Gross, H.J. and Kleinschmidt, A.K. (1976) Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. USA*, **73**, 3852–3856.

Soto, A.M., Misra, V. and Draper, D.E. (2007) Tertiary structure of an RNA pseudoknot is stabilized by 'diffuse' Mg_{2+} ions. *Biochemistry*, **46**, 2973–2983.

Steger, G., Hofmann, H., Fortsch, J., Gross, H.J., Randles, J.W., Sanger, H.L. and Riesner, D. (1984) Conformational transitions in viroids and virusoids: comparison of results from energy minimization algorithm and from experimental data. *J. Biomol. Struct. Dyn.* **2**, 543–571.

Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C. and Weeks, K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **6**, e96.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Compilation of the RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) results for the PLMVd.034 variant of both polarities. For both polarities, the stems already identified are indicated at the bottom of the graph. Nucleotides 262–293 of the (+) strand and 264–293 of the (–) strand are located in a region without results because they are complementary to the first oligonucleotide used in the reverse transcription reaction. The reactions performed with and without magnesium are represented by the black and red bars, respectively. PLMVd, *Peach latent mosaic viroid*.

Fig. S2 Compilation of the RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) results for the PLMVd.282 variant of both polarities. For both polarities, the stems already identified are indicated at the bottom of the graph. Nucleotides 260–291 of the (+) strand and 263–291 of the (–) strand are located in a region without results because they are complementary to the first oligonucleotide used in the reverse transcription reaction. The reactions performed with and without magnesium are represented by the black and red bars, respectively. PLMVd, *Peach latent mosaic viroid*.

Fig. S3 Compilation of the results obtained from RNAhybrid for 306 natural *Peach latent mosaic viroid* (PLMVd) variants related to the formation of the P12 pseudoknot. The sequence loops L1 and L11 of all available PLMVd sequences of (+) polarity that do not have the insertion specific for peach calico disease were analysed using RNAhybrid and are classified according to their respective minimal free energies of hybridization (MFE, kcal/mol) from the most stable to the least stable interaction. The sequences of each loop used for the analysis are indicated for each variant. In each case, we have identified the nucleotides involved in the interaction of the predicted pseudoknot under the columns 'L1 on' and 'L11 on'. The columns 'L1 off' and 'L11 off' indicate the nucleotides that are not implicated in the pseudoknot. The variants that were selected for this study are identified by grey shading.

Fig. S4 Oligonucleotides used for the small *Peach latent mosaic viroid* (PLMVd) molecules.

Fig. S5 Oligonucleotides used for the primer extension reactions performed during the RNA-selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) experiments.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

