This is an author manuscript that has been accepted for publication in *Journal of General Virology*, copyright Society for General Microbiology, but has not been copy-edited, formatted or proofed. Cite this article as appearing in *Journal of General Virology*. This version of the manuscript may not be duplicated or reproduced, other than for personal use or within the rule of 'Fair Use of Copyrighted Materials' (section 17, Title 17, US Code), without permission from the copyright owner, Society for General Microbiology. The Society for General Microbiology disclaims any responsibility or liability for errors or omissions in this version of the manuscript or in any version derived from it by any other parties. The final copy-edited, published article, which is the version of record, can be found at http://vir.sgmjournals.org, and is freely available without a subscription 12 months after publication.

Use of random-mutagenized genomic cDNA banks of *Potato spindle tuber viroid* to screen for viable versions of the viroid genome

Aneta Więsyk[1], Thierry Candresse[2], Włodzimierz Zagórski[1] and Anna Góra-Sochacka[1*]

1: Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5A, 02-106 Warsaw, Poland

2: Equipe de Virologie, UMR GDPP, INRA and Université Bordeaux 2, IBVM, Campus INRA, BP 81, 33883 Villenave d'Ornon cedex, France

Running title: PSTVd *in vivo* Selex for TL and PM1 variants

Key words: Viroid, PSTVd; terminal hairpin; polypurine; pre-melting loop 1; Selex

*: corresponding author.
Phone: +48 22 594 34 08
Fax: +48 22 658 46 36
e-mail annag@ibb.waw.pl

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are GQ853457-GQ853469

Standard (full-length) paper: summary 205 words; main text 4995words, tables and figures 7

SUMMARY

In an effort to study sequence space allowing the recovery of viable *Potato spindle tuber viroid* (PSTVd) variants we have developed an *in vivo* selection (Selex) method to produce and bulk-inoculate by agroinfiltration large PSTVd cDNA banks in which a short stretch of the genome is mutagenized to saturation. This technique was applied to two highly conserved six nucleotide-long regions of the PSTVd genome, the left terminal loop (TL bank) and part of the polypurine stretch in the upper strand of pre-melting loop 1 (PM1 bank). In each case, PSTVd accumulation was observed in a large fraction of bank-inoculated tomato plants. Characterization of the progeny molecules showed the recovery of the parental PSTVd sequence in 89% (TL bank) and 18% (PM1 bank) of the analyzed plants. In addition, viable and genetically stable PSTVd variants with mutations outside of the known natural variability of PSTVd were recovered in both cases, although at different rates. In the case of the TL region, mutations were recovered at five of the six mutagenized positions (357, 358, 359, 1 and 3 of the genome) while for the PM1 regions mutations were recovered at all six targeted positions (50-55), providing significant new insight on the plasticity of the PSTVd genome.

Supplementary material: two supplementary Figures**,** Fig S1.ppt & Fig S2.ppt

INTRODUCTION

Viroids are small, single-stranded, circular, non-encapsidated and non-coding RNA molecules causing diseases in various plants (Hadidi *et al.,* 2003; Flores *et al.,* 2005; Tsagris *et al.,* 2008) that have been classified in two families, the *Avsunviroidae* and the *Pospiviroidae* (Flores *et al.,* 2003). The best known representative of the *Pospiviroidae* family is *Potato spindle tuber viroid* (PSTVd), which typifies the *Pospiviroid* genus. Based on sequence comparison, the *Pospiviroid* genome rod-like structure has been divided into five structural domains named central (C), pathogenicity (P), variable (V) and terminal right (TR) and left (TL) (Keese & Symons, 1985) (Fig. 1). The C domain together with the TL domain is involved in replication (Baumstark *et. al*., 1997; Schrader *et al.,* 2003; Kolonko *et al.,* 2006; Zhong *et al.,* 2008) whereas the P domain seems to control pathogenicity (Schnölzer *et al.,* 1985; Owens *et al.,* 1996). Structural elements of the TR and V domains are important for viroid trafficking in host plants (Qi *et al.,* 2004; Zhong *et al.,* 2007). Since viroids do not encode proteins, they depend on host enzymes for their biological functions. It is now clear that most viroid functions are collectively regulated by determinants located in different parts of the viroid molecule (Zhong *et al.,* 2008).

Site-directed mutagenesis studies have allowed the unraveling of particular PSTVd structures and their contribution to PSTVd biology (Owens *et al.,* 1995, 1996; Hu *et al.,* 1996, 1997; Qi *et al.,* 2004; Wang *et al.,* 2007; Zhong *et al.,* 2008). However, over 50% of randomly introduced mutations abolish PSTVd infectivity (Owens *et al.,* 1991) and mutations are frequently unstable (Owens *et al.,* 1995, 1996), limiting the usefulness of this strategy to probe in detail the sequence space around a particular viroid sequence.

An alternative strategy is the *in vivo* selection (*in vivo* Selex) of viable variants from large collections of mutants (Owens *et al.,* 2003; Owens & Thompson, 2005). Starting from libraries of mutants with genomes partially or fully randomized at key positions, these authors

inoculated tomato plants with *in vitro* transcribed PSTVd RNAs and analyzed the recovered progenies, identifying a number of new, viable variants. Another strategy based on biolistic inoculation was used to screen viroid "thermomutant" propagation in different plant species (Matousek *et al.*, 2004a, 2004b). In the present work we have developed and used an *in vivo* Selex approach to study the viable variability in two highly conserved 6 nucleotide-long segments of the PSTVd molecule located respectively in the left terminal loop of the TL domain and in the upper strand of pre-melting loop 1 (PM1) of the P domain. The left terminal hairpin is highly conserved between natural PSTVd isolates and in all other *Pospiviroids* characterized to date. A search of over 140 PSTVd sequences present in GenBank revealed only four natural isolates diverging from the $C_{357}UUCGG_3$ motif. These isolates, AY532801-804 are also unusual in having been identified from pepper plants (Lebas *et al.*, 2005), a rare natural host for PSTVd.

The second targeted region is part of the so-called oligo- or polypurine stretch (Steger *et al.*, 1984), a hallmark of the P domain of *Pospiviroids* (Keese & Symons, 1985). In PSTVd, the consensus of this domain is $AGA_{(3-5)}GA_{(5-7)}GA$ with only two sequences in GenBank for which a pyrimidine is present in this purine stretch. Such a purine-rich motif is also observed in many other *Pospiviroidae*. Despite this conservation, no clear biological function has been associated with this motif, beyond its contribution to the local instability of the molecule (Steger *et al.*, 1984; Gast *et al.*, 1996). Mutations immediately upstream of this region or on the corresponding lower strand of the PSTVd molecule affect the pathogenicity (Schnölzer *et al.*, 1985; Hammond & Owens, 1987; Owens *et al.*, 1996) or the trafficking of PSTVd isolates (Qi *et al.*, 2004; Zhong *et al.*, 2007).

The methods developed allowed the preparation and screening of banks containing all possible mutants of the six nucleotide target regions. Sequencing of the PSTVd progenies recovered from the inoculated plants allowed the identification of viable mutants from pools of variants obtained by saturation mutagenesis.

RESULTS

In vivo selection of viable PSTVd variants from the PM1 and TL mutant banks

Following the PCR-based strategies described in the Methods section, two *Agrobacterium tumefaciens* banks containing mutagenized full-length monomeric infectious PSTVd constructs under the control of the CaMV 35S promoter in the pGreen binary vector (Hellens *et al.,* 2000) were prepared. Both banks correspond to the saturation mutagenesis of a short 6 nucleotide-long region corresponding respectively to the left terminal hairpin (TL, positions 357-359 and 1-3 of the PSTVd molecule, numbering according to the PSTVd-DI reference isolate sequence) and to part of the upper portion of pre-melting loop 1 (PM1) in the virulence modulating (VM; Schnölzer *et al.,* 1985) region (positions 50-55) (Fig. 1). Each bank contained over 50,000 individual colonies and therefore contains with high statistical probability all 4096 (4^6) possible PSTVd variants. Five clones of each bank were chosen randomly and sequenced, confirming the presence of mutations in the target regions. No additional mutations outside the mutagenized regions were observed (result not shown).

Each bank was bulked, amplified and mass-agroinoculated to the leaves of 50 tomato plants cv. 'Rutgers'. As a control, 10 plants were inoculated with *A. tumefaciens* carrying the parental PSTVd-S23 cDNA similarly cloned in the pGreen vector. Two weeks after agroinfiltration, PSTVd accumulation was detected in 60% of the control PSTVd-S23 inoculated plants and this proportion reached 100% by 4 weeks post inoculation (wpi) (Table 1), thus confirming the efficiency of the inoculation method. By contrast, only 58% (29/50) and 24% (12/50) of plants inoculated with the TL and PM1 banks, respectively, showed PSTVd accumulation by 4 wpi (Table 1) but in both cases this proportion reached 92% by 8 wpi. The hybridization signals observed with the bank-inoculated plants were generally lower than those of PSTVd-S23 inoculated controls, but showed a general tendency to increase with

time, so that a significant portion of plants ultimately reached PSTVd accumulation levels similar to those of the control plants (Supplementary Fig. S1A and B).

Analysis of the TL bank progeny

The PSTVd progeny present in 18 of the 46 plants inoculated with the TL bank were characterized by direct sequencing of PCR-amplified full-length cDNAs (Góra *et al.*, 1994). In 16 of these 18 plants only the PSTVd-S23 parent was observed and these plants were not analyzed further. There was no clear correlation between the onset of detectable PSTVd accumulation and detection of the PSTVd-S23 sequence since the first positive hybridization signal for these 16 plants was observed at between 2 and 6 wpi. The two other plants analyzed yielded mutants differing from the PSTVd-S23 sequence at 3 or 4 of the 6 mutagenized positions. No mutations were observed in other parts of the genome of the recovered progenies. Plant number 34 analyzed at 3 wpi accumulated variant TL34/3, which differs from PSTVd-S23 by three mutations in the terminal loop: C358A, U359A and C1A (Table 2 and Fig. 2A). The same variant was observed again in the same plant at 4 wpi but later analysis, at 7 wpi, revealed the accumulation of another variant, TL34/7, differing from TL34/3 by a A358G mutation. No sequence heterogeneity was observed at any of these three time points. No symptoms were observed at 3 wpi on plant 34 and the hybridization signal was very weak but by 4 wpi the hybridization signal was comparable to that of control S23-inoculated plants and intermediate severity symptoms were observed. At 7 wpi the plant displayed typical S23 severe symptoms (results not shown).

A third variant, TL4/11, was also detected in a single plant, differing from the S23 parent by 4 mutations: C357A, U359A, C1A and G3C (Table 2 and Fig. 2A). This mutant was detected at 5 wpi in a plant showing mild symptoms and a strong hybridization signal but was not further analyzed.

Collectively, the three recovered variants differ from the parental sequence at five of the 6 mutagenized positions, the G at position 2 being the only parental nucleotide retained in all three variants which, in addition, share the same pair of U359A and C1A mutations (Table 2).

The infectivity and genetic stability of these three variants were verified by inoculating infectious monomeric cDNAs to tomato seedlings and resequencing their progeny at 4 wpi. The results (Table 2) confirm the infectivity of all three variants. TL34/3 and TL34/7 induced severe symptoms similar to those of PSTVd-S23 while TL4/11 only caused mild symptoms (result not shown). All three variants were genetically stable since in each case a sequence identical to that of the inoculated variant was obtained when analyzing its progeny at 4 wpi (result not shown). However, all three variants also showed a somewhat reduced accumulation as compared to the S23 parent. Although this effect was rather limited in the case of TL34/3 and 34/7, it was stronger for TL4/11 (Fig. 3).

Analysis of the PM1 bank progeny

The PSTVd progeny was characterized in 33 of the 46 plants infected upon inoculation with the PM1 bank. In five of these plants the parental PSTVd-S23 was detected. As for the TL bank, S23 was observed in plants showing indifferently early (4 wpi) or late (up to 8 wpi) onset of PSTVd accumulation. The other 28 plants yielded a total of 10 variants differing by between 1 and 5 mutations from the parental sequence. Overall, all 6 of the mutagenized positions were observed to be affected in one or more variants (Table 3). While three of the variants contained additional point mutations outside of the mutagenized region, none of them contained insertions or deletions. Six of the 11 variants were observed in more than one plant and one third of the plants analyzed (11/33) contained more than a single sequence variant (Table 3).

Two of the recovered variants are characterized by single mutations affecting position 50. In plant number 16, variant PSTVd-PM1-16, carrying a single U50C mutation, was observed at 6 wpi. However, two weeks later the population was heterogeneous. Sequencing of six cloned PSTVd cDNAs revealed the presence of three variants, PSTVd-PM1-16 (one clone), the parental PSTVd-S23 (four clones) and a new variant, PSTVd-PM1-46 (one clone). The last one differs from PSTVd-PM1-16 and PSTVd-S23 by the presence of an A at position 50. As compared to the parental PSTVd-S23, this U50A substitution corresponds to reversion to the PSTVd-DI sequence (Gross *et al.*, 1978). Mutant PM1-46 was also detected, alone, in another plant (plant 46, Table 3).

A second group of five variants is characterized by a pair of G54A-A55G mutations. These variants are by far the most frequently recovered, since they were collectively observed in 82% of the plants containing non-parental sequences (23 of 28, Table 3). Mutant PSTVd-PM1-40 carries only these two mutations and was detected at 6 wpi in 3 plants (11, 21 and 40). In the case of plant 40 heterogeneity at position 50 was observed and sequencing of cloned PSTVd cDNAs revealed four clones with an A50 and one clone with an U50. Interestingly, the variant with the additional A50U mutation, named PSTVd-PM1-9 was the most frequently observed molecule, since it was also detected as the only variant present in nine plants (plants 1, 2, 4, 8, 9, 15, 25, 44 and 50) and, in mixed infection with other variants, in nine other plants (plants 11, 17, 20, 21, 24, 30, 37, 41 and 47) (Table 3). The variant PSTVd-PM1-31 which differs from PM1-40 and PM1-9 by a single U50C mutation, was detected in a single plant (31).

In three of the plants (20, 24, 30), PSTVd-PM1-9 was observed in co-infection with a variant PSTVd-PM1-30, carrying an additional A51C mutation. In all three cases, this variant showed a tendency to be displaced by PM1-9. For plant 30, PM1-30 was observed alone 3 wpi but 2 weeks later was observed in mixed infection with PSTVd-PM1-9. The same situation was observed for plant 24 at 6 and 8 wpi. In the case of plant 20, sequence analysis

of the uncloned RT-PCR products at 5 wpi revealed the presence of the two variants but reanalysis three weeks later yielded only PSTVd-PM1-9. The presence of an A at position 51 seems to be favored over a C. However, PSTVd-PM1-30 was also observed alone in three additional plants (6, 27 and 48), even as late as 6 wpi in the case of plant 48.

Mixed infection with PSTVd-PM1-9 was similarly observed in four plants (17, 37, 41 and 47) for another variant, PSTVd-PM1-37, which differs from PM1-9 by an additional A52C substitution. Variant PM1-37 was never observed alone.

Lastly, three variants containing additional mutations outside the mutagenized region were observed alone in single plants and late in the experiment (6-7 wpi). In comparison with the parental S23 sequence, variant PSTVd-PM1-14 has mutation in 5 of the six targeted positions: U50C, A52U, A53G, A54U and A55G and has an additional G44A mutation.

Variant PSTVd-PM1-43 has three mutations in the target region (U50C, A52U, G54A) plus an additional U306A mutation in the lower strand of the VM region, exactly opposite the mutagenized region. Similarly, variant PSTVd-PM1-7 has two mutations in the target region (A53G and G54A) and an additional U311A mutation in the lower portion of the VM.

The infectivity and genetic stability of all PM1 variants were verified as for the TL mutants. The results obtained (Table 3) confirmed the infectivity of all 10 variants, even if the infectivity of variant PM1-43 seems limited, as it infected only 3 of the 10 inoculated plants. Most variants induced severe symptoms similar to those of the parental S23 isolate but intermediate severity symptoms were observed for variants PM1-9, PM1-30, and PM1-7 while heterogeneity in symptom severity was observed in the case of variants PM1-30, PM1-14 and PM1-43. However, this heterogeneity does not appear to correlate with genetic instability since parallel analysis of progeny molecules from plants with mild or severe infection phenotypes yielded in each case only the sequence of the inoculated parental variant. With the exception of PSTV-PM1-37, all variants were therefore genetically stable over the 5

week study and no evidence for sequence heterogeneity was detected in the progeny sequencing chromatograms.

Mutant PSTVd-PM1-37 displayed high infectivity (20 of 22 inoculated plants) but in three of the 4 plants in which its progeny was analyzed the sequencing chromatograms provided indications of C/A sequence heterogeneity at position 52, which differentiates PM1-37 from PM1-9. The last plant analyzed yielded progeny only composed of PSTVd-PM1-9 (Table 3).

Most variants accumulated to levels comparable to that of the PSTVd-S23 parent but four of them (PM1-14, PM1-30, PM1-31, PM1-43) showed a reduced accumulation (Supplementary Fig. S2). For PM1-30 and PM1-31, later resampling at 6 wpi demonstrated that accumulation levels had caught up with that of S23 (Supplementary Fig. S2).

DISCUSSION

In the present study, highly conserved continuous sequence stretches were targeted and no specific measure was taken to preclude the presence of the parental sequence from the inoculated pool of variants. As a consequence, the complexity of the banks used (4^6=4096 variants) was significantly higher than the 48 to 3072 variant pools previously used (Owens *et al.*, 2003; Owens & Thompson, 2005). It should be noted that while excluding the parental PSTVd sequence from the mutant pool reduces the sampled sequence space, it does not guarantee that the wild type sequence will not be recovered, sometimes as the sole variant, during the *in vivo* Selex step (Owens & Thompson, 2005). In the present study, the wild type PSTVd-S23 sequence was present in the initial TL and PM1 banks and was recovered in very different proportions following the *in vivo* selection, (89% and 18%, respectively). Although the two target regions are highly conserved between PSTVd isolates, this difference in the rate of recovery of the wild type parent probably reflects differences in the paucity of viable

and/or competitive variants in the banks analyzed. The only variants recovered from the TL bank diverged by 3-4 mutations from the S23 parent, whereas the variant most frequently selected from the PM1 bank, PM1-9, had only 2 mutations and some other variants differed from S23 by a single mutation.

Of the over 140 PSTVd sequences in GenBank only four natural isolates have left terminal hairpins diverging from the $C_{357}UUCGG_3$ motif. There are, in addition, three reports of viable and stable mutants affected in this region. One involves a variant with a U359A mutation recovered during an *in vivo* Selex experiment (Owens *et al.,* 2003) and the others involve C1U or C1G point mutants (Owens *et al.,* 1991; Kolonko *et al.,* 2006). On the other hand, a U359G mutant was unstable and reverted to the wild type sequence (Kolonko *et al.,* 2006). Enlargement of the terminal loop from 4 to 6 nucleotides as a result of three mutations (G2U, A4C and C6G), reduced PSTVd replication in *Nicotiana benthamiana* to below 16% of the wild type sequence (Zhong *et al.,* 2008). The same set of mutations completely abolished PSTVd infectivity in tomato (Hammond & Owens, 1987; Hammond, 1994; Hu *et al.,* 1997). However, restoration of replication and systemic trafficking in tomato only required a C4G mutation (Hu *et al.,* 1997).

Two possible structures for the TL domain were proposed, the well known rod-like structure and a branched structure involving two terminal hairpins (Gast *et al.,* 1996; Gast, 2003). NMR and thermodynamic analysis of mutations predicted to affect these conformations indicated that the elongated-rod form was the preferred structure in solution (Dingley *et al.,* 2003). Similarly, analysis of *in vivo* Selex variants led Owens and Thompson (2005) to conclude that the branched conformation was unlikely to be important for PSTVd fitness. The analysis of the potential effects on both conformations of the mutations observed in the three TL variants recovered in the present study similarly do not support a significant role for the branched conformation. Indeed, three of these mutations, at positions 357, 358 and 3 are predicted to affect base-pairing in the bifurcated structure. In particular, the C357A

and G3C mutations observed in variant PSTVd-TL4/11 are predicted to affect a total of 4 base pairs in the branched structure but a single one in the elongated conformation (Fig. 2B).

Some tetraloops may adopt particular tertiary structures involving non-canonical base pairings that significantly contribute to the stabilization of the molecule (Moore, 1999). The PSTVd cCUCGg terminal loop may belong to the CUNG or YNMG tetraloop families (Proctor *et al.,* 2002; Kolonko *et al.,* 2006). In model experiments, a CUCG loop with a closing C:G nucleotide pair had a high thermodynamic stability (Proctor *et al.,* 2002). However, the mutations observed in the present study are all expected to significantly reduce loop stability (ΔTm of between -3.6 to -6.3 °C, $\Delta\Delta G_{37°}$ of +1.45 to +4.74 kcal/mol, Proctor *et al.,* 2002), Although stabilization of the PSTVd molecule through such a tetraloop effect might explain the extreme conservation of the terminal loop sequence in *Pospiviroids*, it appears that this stabilization is not a prerequisite for PSTVd infectivity or stability in tomato plants.

The biological function(s) of the TL domain is (are) not entirely understood but a contribution to the PSTVd molecule stability (Hammond & Owens, 1987; Hammond, 1994; Hu *et al.,* 1997) and replication process has been demonstrated. Kolonko *et al.* (2006) showed that the synthesis of (-) strand is initiated at either position U_{359} or C_1. Site-directed mutagenesis indicated that a C1G mutation was viable and stable but a U359G mutation rapidly reverted to the wild type sequence (Kolonko *et al.,* 2006), suggesting U359 as the most probable transcription start site (TSS).

The mutants reported here indicate that both mutations U359A and C1A are viable and stable. Mutations affecting the other mutagenized positions were also recovered, with the exception of position 2, for which the parental G was always observed. Both strands of the PSTVd genome are replicated by the host DNA-dependent RNA polymerase II (Pol II, Schindler & Mühlbach, 1992). Pol II promoters show extensive variability, even if a -1 Y and +1 R (+1 being the TSS) nucleotide dimer sequence seems to be significantly over-

represented in plant Pol II promoters (Yamamoto *et al.,* 2007; Shamuradov *et al.,* 2003). In this context it may be noteworthy that PSTVd seems to tolerate stably any nucleotide at position 1, a C in the wild type sequence, A in the mutants recovered in the present study and a G or a U in previously observed or constructed mutants (Owens *et al.,* 1991; Kolonko *et al.,* 2006). By contrast, position 359 seems to be somewhat more constrained as a U359G mutation proved unstable (Kolonko *et al.,* 2006) and only an A was recovered at this position in the present work. Taken together these results suggest that U359 could represent the TSS during PSTVd (-) strand transcription.

The second targeted region, in the polypurine stretch of pre-melting loop 1 (Steger *et al.,* 1984) is also almost absolutely conserved in natural PSTVd isolates and, when analyzed, mutations introduced in this region have proven unstable (Owens, 1990; Zhong *et al.,* 2008).

The rates of recovery of PM1 variants showed an almost 20-fold variation, ranging from 57% (19 of 33 plants) to 3% (1 of 33 plants). Since the strategy used should have ensured an unbiased representation of all possible mutants in the inoculum, the variation observed in recovery rate most likely reflects the competitiveness of the individual variants. However, it is not possible to directly assimilate the recovery rate with genetic stability, as demonstrated by the recovery of the stable parental PSTVd-S23 (Góra *et al.,* 1997) and of the unstable variant PM1-37 at roughly similar rates (18% and 12%, respectively). It seems however noteworthy that with the exception of PM1-37, all the PM1 variants recovered proved stable over a 5 week assay period, whereas previous limited mutagenesis efforts in that region had yielded only unstable molecules (Owens, 1990; Zhong *et al.,* 2008).

It is remarkable that molecules in which one or more of the target bases was replaced by a pyrimidine represented 9 of the 11 recovered variants, including the parental PSTVd-S23 sequence. In fact, some of the pyrimidine containing variants, such as PSTVd-S23 and PM1-30, were among the most frequently recovered molecules. While mutations were recovered at every target position, purine to pyrimidine mutations were observed at positions 50, 51, 52,

and 54. Despite the extremely high conservation of the polypurine stretch in *Pospivoiroids*, the results reported here demonstrate that PSTVd isolates with pyrimidines in one or more of the target positions can be viable, genetically stable and have significant fitness, as judged from their recovery rates and accumulation levels.

The two recovered variants without pyrimidines are PM1-46 and PM1-9. Their rate of recovery (6% and 57%, respectively) are not correlated with the frequency of natural PSTVd isolates with the same local sequence. Indeed, isolates with the AAAAGA sequence found in PM1-46 represent 60% of the PSTVd sequences in GenBank while those with AAAAAG, as PM1-9, represent a mere 2% (3 isolates). We have no explanation for this observation, in particular since PM1-9 and PM1-46 accumulated to comparable levels (Supplementary Fig. S2).

If one considers the range of mutants present in the initial inoculum, the potential local secondary structures vary from a 50-55 region completely base paired to positions 305-310 of the lower strand to a completely unpaired region (Fig. 4). The corresponding folding ΔG values for the complete PSTVd molecule calculated using Mfold 3.2 (Zuker, 2003) range from -165.7 to -156.5 kcal/mol. Similar computations provide a -159.5 kcal/mol for the parental S23 sequence and a range of -158.4 (PM1-9 and PM1-37) to -162.3 (PM1-7) kcal/mol for the recovered variants indicating that they only represent a central fraction of the secondary structure stability levels. It is noteworthy that for two of the mutants with additional mutations, PM1-7 and PM1-14, these mutations lower the ΔG value (0.4-0.5 kcal/mol), suggesting that the most stable structures/mutants could have been counter-selected (Fig. 4). Such an observation would be in keeping with results showing that complete base pairing of the target region though pairs of mutations affecting either positions 50-51 or 310-311 resulted in mutants that were either noninfectious (Owens *et al.,* 1996) or severely affected in their replication and unstable (Zhong *et al.,* 2008). Taken together, the results reported here suggest that viability, genetic stability and fitness of PSTVd molecule can be

achieved even if pyrimidines are introduced in the polypurine stretch, provided that the local secondary structure remains in a limited range of stability, with mutations either stabilizing or further destabilizing the region showing detrimental effects.

Although it is unlikely that all possible viable variants in the target mutagenized regions were recovered in the experiments reported here, the approach used allowed the identification of numerous genetically stable variants that fall outside the natural variability of PSTVd and provide new insights into the plasticity of the PSTVd genome. Further efforts employing the same strategy, in particular with larger target regions or targeting other genomic regions should improve our knowledge of the viable sequence space surrounding the genomes of viroids.

METHODS

Construction of the PSTVd-PM1 bank.

Random saturation mutagenesis of positions 50-55 of a full-length infectious PSTVd-S23 (Góra *et al.,* 1994) cDNA was achieved by PCR amplification using two primers, one of which was fully degenerate at the 6 target positions. Primer, PSTVd-P1 (5'G**GGATCC**CCGGGGAAACCT3', *Bam*HI site of the central conserved region in **bold**) corresponds to PSTVd positions 86-104. Primer PSTVd-P4mut (5'G**GGATCC**CTGAAGCGCTCCTCCGAGCCGCCTTCTTTTTNNNNNNCTGCTCAGG AGGTCAGGTGT3') is complementary to PSTVd positions 93-30. The amplification reaction was performed using *Pfu* DNA polymerase (Fermentas) and the following cycling scheme: 30 sec at 94 °C , 30 sec at 64 °C and 1 min at 74 °C, for 31 cycles. The PCR product was cloned into pGEM-T-Easy (Promega), generating a bank of mutagenized PSTVd monomers. Recombinant plasmids were purified from the bulked bacterial colonies (Qiagen plasmid purification kit), the full-length PSTVd inserts excised using *Bam*HI and subcloned into the corresponding site of a pGreen vector (Hellens *et al.,* 2000) modified by replacing the

original polylinker by that of pUC9. Depending on the orientation of the viroid cDNA, this strategy generates an 8 or 11 bp duplication of the central conserved region of the PSTVd genome which is sufficient for infectivity (Candresse *et al.,* 1990). Following transformation of *Agrobacterium tumefaciens* [C58C1 strain carrying the virulence helper plasmid pCH32 (Hamilton *et al.,* 1996)] with a pool of the recombinant pGreen vectors, a bank of over 50,000 bacterial colonies was obtained.

Construction of the PSTVd-TL bank

Two terminally overlapping PSTVd-S23 PCR products were first obtained using two different primer pairs. The first PCR product covers positions 86-24 of the circular genome and was obtained using primers TL1 (5'G**GGATCC**CCGGGGAAACCT3', PSTVd positions 86-104, *Bam*HI site of the central conserved region in **bold)** and TL5 (5'CAGGAACCACGAGTTTAGTTNNNNNNAACCAACTGCGGTTCCAAGG3', PSTVd positions 24-1/359-336, with fully randomized positions 357-359 and 1-3 indicated by Ns). The second PCR product, spanning positions 4-93 was obtained using primers TL3 (5'AACTAAACTCGTGGTTCCTGT3', positions 4-24) and primer TL2 (5'G**GGATCC**CTGAAGCGCTCC3', complementary to positions 74-93, *Bam*HI site in **bold)**. The two PCR products were purified and joined into full-length PSTVd monomeric cDNAs by a fusion-PCR using primers TL1 and TL2 for this second round of amplification. All PCR reactions were carried out using *Pfu* polymerase and all further steps for the preparation of an agroinfiltrable bank were carried out as described above for the PM1 bank.

Mass inoculation of tomato plants and recovery of infectious PSTVd progeny molecules

One month-old tomato plants cv. 'Rutgers' were inoculated by agroinfiltration of the bulked TL or PM1 banks as described by Bendahmane *et al.* (2000). From two to eight weeks after inoculation plants were regularly observed for symptom development, leaf samples were

collected and the presence of PSTVd assessed by dot-blot hybridization. Total RNAs were extracted from individual plants with the RNase Plant mini kit (Qiagen) and subjected to RT-PCR using PSTVd-specific primers complementary to the CCR region (Góra *et al.*, 1994). Uncloned PCR products were directly sequenced and chromatograms checked for signs of sequence heterogeneity. Individual monomeric infectious cDNAs of progeny molecules cloned in plasmid pUC9 were obtained as described previously (Góra *et al.*, 1994) and their sequences verified.

Verification of the infectivity and genetic stability of progeny molecules

Tomato cv. 'Rutgers' seedlings were inoculated as described previously (Candresse *et al.*, 1990) with purified plasmids (2 μg/plant) containing monomeric full-length cDNAs of all PSTVd progeny variants. The inoculated plants were monitored for symptom development, their infection status determined by molecular hybridization and the genetic stability of the PSTVd variants determined by direct sequencing of PCR-amplified progeny PSTVd cDNAs as described above.

Secondary structure and stability predictions for the isolated variants

Secondary structures and folding stability predictions at 37 °C in 1 M NaCl were performed for all recovered variants using the Mfold 3.2 program (Zuker, 2003).

ACKNOWLEDGEMENTS

We gratefully thank Anne-Lise Haenni for reviewing the English of the manuscript.

REFERENCES

Baumstark, T. Schröder, A. & Riesner, D. (1997). Viroid processing: switch form cleavage to ligation is driven by a change from a tetraloop to a loop E conformation. *EMBO J.* **16,** 599-610.

Bendahmane, A., Querci, M., Kanyuka, K. & Baulcombe, D.C. (2000). *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: Application to the Rx2 locus in potato. *Plant J.* **21,** 73-81.

Candresse, T., Diener, T.O. & Owens, R.A. (1990). The role of the viroid central conserved region in cDNA infectivity. *Virology* **175,** 232-237.

Dingley, A.J., Steger, G., Esters, B., Riesner, D. & Grzesiek, S. (2003). Structural characterization of the 69 nucleotide potato spindle tuber viroid left-terminal domain by NMR and thermodynamic analysis. *J. Mol. Biol*. **334,** 751-767.

Flores, R., Hernández, C., Martínez de Alba, A.E., Daròs, J.A. & Di Serio, F. (2005). Viroids and viroid-host interactions. *Ann. Rev. Phytopath*. **43,** 117-139.

Flores, R., Randles, J.W. & Owens, R.A. (2003). Classification. In *Viroids*, pp.71-75. Edited by A. Hadidi, R. Flores, R.W. Randles, & J.S. Semancik. CSIRO Publishing, Collingwood, VIC, Australia.

Gast, F.U. (2003). A new structural motif in the left terminal domain of large viroids identified by covariation analysis. *Virus Genes* **26,** 19-23.

Gast, F.U., Kempe, D., Spieker, R.L. & Sänger, H.L. (1996). Secondary structure probing of potato spindle tuber viroid (PSTVd) and sequence comparison with other small pathogenic RNA replicons provides evidence for central non-canonical base-pairs, large A-rich loops, and a terminal branch. *J. Mol. Biol*. **262,** 652-670.

Góra, A., Candresse, T. & Zagórski, W. (1994). Analysis of the population structure of three phenotypically different PSTVd isolates. *Arch. Virol*. **138,** 233-245.

Góra, A., Kierzek, A., Candresse, T. & Zagórski, W. (1997). The genetic stability of potato spindle tuber viroid (PSTVd) molecular variants. *RNA* **3,** 68-74.

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

Hadidi, A., Flores, R., Randles, R.W. & Semancik, J.S. (2003). *Viroids*. CSIRO Publishing, Collingwood, VIC, Australia.

Hamilton, C.M., Frary, A., Lewis, C. & Tanksley, S.D. (1996). Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc. Natl. Acad. Sci. USA* **93,** 9975-9979.

Hammond, R.W. (1994). *Agrobacterium*-mediated inoculation of PSTVd cDNAs onto tomato reveals the biological effects of apparently lethal mutations. *Virology* **201,** 36-45.

Hammond, R.W. & Owens, R.A. (1987). Mutational analysis of potato spindle tuber viroid reveals complex relationships between structure and infectivity. *Proc. Natl. Acad. Sci. USA* **84,** 3967-3971.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S. & Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* **42,** 819-832.

Hu, Y., Feldstein, P.A., Bottino, P.J. & Owens, R.A. (1996). Role of the variable domain in modulating potato spindle tuber viroid replication. *Virology* **219,** 45-56.

Hu, Y., Feldstein, P.A., Hammond, J., Hammond, R.W., Bottino, P.J. & Owens, R.A. (1997). Destabilization of potato spindle tuber viroid by mutations in the left terminal loop. *J. Gen. Virol*. **78,** 1199-1206.

Keese, P. & Symons, R.H. (1985). Domains in viroids : evidence for intermolecular RNA rearrangements and their contribution to viroid evolution. *Proc. Natl. Acad, Sci. USA* **82,** 4582-4586.

Kolonko, N., Bannach, O., Aschermann, K., Hu, K.H., Moors, M., Schmitz, M., Steger, G. & Riesner, D. (2006). Transcription of potato spindle tuber viroid by RNA polymerase II starts in the left terminal loop. *Virology* **347,** 392-404.

Lebas, B.S.M., Clover, G.R.G., Ochoa-Corona, F.M., Elliott, D.R., Tang, Z., & Alexander, B.J.R. (2005). Distribution of potato spindle tuber viroid in New Zealand glasshouse crops of *Capsicum* and tomato. *Austr. Plant Path.* **34,** 129-133.

Matoušek, J., Orctová, L., Steger, G., Škopek, J., Moors, M., Dědič, P. & Riesner, D. (2004a). Analysis of thermal stress-mediated PSTVd variation and biolistic inoculation of progeny of viroid "thermomutants" to tomato and *Brassica* species. *Virology* **323,** 9-23.

Matoušek, J., Orctová, L., Steger, G. & Riesner, D. (2004b). Biolistic inoculation of plants with viroid nucleic acids. *J. Virol. Meth* **122,** 153-164.

Moore, P.B. (1999). Structural motifs in RNA. *Annual Review of Biochemistry* **68,** 287-300.

Owens, R.A. (1990). Mutational analysis of viroid pathogenicity: Tomato apical stunt viroid. *Molec. Plant Microbes Interact.* **6,** 374-380.

Owens, R.A., Chen, W., Hu, Y. & Hsu, Y.H. (1995). Suppression of potato spindle tuber viroid replication and symptom expression by mutations which stabilize the pathogenicity domain. *Virology* **208,** 554-564.

Owens, R.A., Steger, G., Hu, Y., Fels, A., Hammond, R.W. & Riesner, D. (1996). RNA structural features responsible for potato spindle tuber viroid pathogenicity. *Virology* **222,** 144-158.

Owens, R.A. & Thompson, S.M. (2005). Mutational analysis does not support the existence of a putative tertiary structural element in the left terminal domain of Potato spindle tuber viroid. *J. Gen. Virol.* **86,** 1835-1839.

Owens, R.A., Thompson, S.M. & Kramer, M. (2003). Identification of neutral mutants surrounding two naturally occurring variants of Potato spindle tuber viroid. *J. Gen. Virol.* **84,** 751-786.

Owens, R.A., Thompson, S.M. & Steger, G. (1991). Effects of random mutagenesis upon potato spindle tuber viroid replication and symptom expression. *Virology* **185,** 18-31.

Proctor, D.J., Schaak, J.E., Bevilacqua, J.M., Falzone, C.J. & Bevilacqua, P.C. (2002). Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry* **41,** 12062-12075.

Qi, Y., Pélissier, T., Itaya, A., Hunt, E., Wassenegger, M. & Ding, B. (2004). Direct role of a viroid RNA motif in mediating directional RNA trafficking across a specific cellular boundary. *Plant Cell* **16,** 1741-1752.

Schindler, I.M. & Mühlbach, H.P. (1992). Involvement of nuclear DNA-dependent RNA polymerases in potato spindle tuber viroid replication: a reevaluation. *Plant Sci*. **84,** 221-229.

Schnölzer, M., Haas, B., Raam, K., Hofmann, H. & Sänger, H.L. (1985). Correlation between structure and pathogenicity of potato spindle tuber viroid (PSTV). *EMBO J.* **4,** 2181-2190.

Schrader, O., Baumstark, T. & Riesner, D. (2003). A mini-RNA containing the tetraloop, wobble-pair and loopE motifs of the central conserved region of potato spindle tuber viroid is processed into a minicircle. *Nucleic Acids Res.* **31,** 988-998.

Shamuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M. & Solovyev, V.V. (2003). PlantProm: a database of plant promoter sequence. *Nucleic Acids Res*. **31,** 114-117.

Steger, G., Hofmann, H., Förtsch, J., Gross, H.J., Randles, J.W., Sänger, H.L. & 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.

Tsagris, E.M., Martinez de Alba, A.E., Gozmanova, M., Kalantidis, K. (2008). Viroids. *Cell Microbiol.* **10,** 2168-79.

Wang, Y., Zhong, X., Itaya, A. & Ding, B. (2007). Evidence for the existence of the loop E motif of Potato spindle tuber viroid *in vivo*. *J. Virol*. **81,** 2074-2077.

Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K. & Abe, T. (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* **8,** 67.

Zhong, X., Tao, X., Stombaugh, J., Leontis, N. & Ding, B. (2007). Tertiary structure and function of an RNA motif required for plant vascular entry to initiate systemic trafficking. *EMBO J.* **26,** 3836-3846.

Zhong, X., Archual, A.J., Amin, A.A. & Ding, B. (2008). A genomic map of viroid RNA motifs critical for replication and systemic trafficking. *Plant Cell* **20,** 35-47.

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

Table 1. Number of plants showing PSTVd accumulation in the upper non-agro-inoculated parts of tomato cv 'Rutgers' plants

Agroinoculated construct	Time post-inoculation						
	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks	7 weeks	8 weeks
PSTVd-S23	6/10*	8/10	10/10	10/10	10/10	10/10	10/10
Bank TL	2/50	18/50	29/50	35/50	44/50	44/50	46/50
Bank PM1	0/50	2/50	12/50	30/50	37/50	43/50	46/50

*The values given represent the number of plants showing PSTVd accumulation over the total number of agro-inoculated plants.

Table 2. Name, sequence at the targeted positions, infectivity and symptomatology of the PSTVd variants recovered from tomato plants agro-inoculated with TL bank

PSTVd variant	Targeted genome positions*						Infectivity†	Symptoms‡
	357	358	359	1	2	3		
PSTVd S23	C	C	U	C	G	G	5/5	Severe
PSTVd-TL34/3	C	**A**	**A**	**A**	G	G	9/12	Severe
PSTVd-TL34/7	C	**G**	**A**	**A**	G	G	11/12	Severe
PSTVd-TL4/11	**A**	C	**A**	**A**	G	**C**	25/27	Mild

*Numbering according the the PSTVd-DI reference isolate sequence. Nucleotides differing from the parental PSTVd-S23 sequence are indicated in bold.

†The values given represent the number of plants showing PSTVd accumulation over the total number of tomato plants inoculated with plasmids harboring a full-length infectious cDNA of the relevant variant.

‡Symptoms observed at 5 wpi on the tomato plants inoculated with plasmids harboring a full-length infectious cDNA of the relevant variant.

Table 3. Name, sequence at the targeted positions, additional mutations observed, infectivity and symptomatology of the PSTVd variants recovered from tomato plants agro-inoculated with PM1 bank

PSTVd variants	Targeted genome positions*						Additional mutations	Observed in plant number†	Infectivity‡	Symptoms§
	50	51	52	53	54	55				
S23	U	A	A	A	G	A	-	16, 18, 22, 26, 33, 40, 42	5/5	Severe
PM1-7	U	A	A	**G**	**A**	A	U311A	7	9/10	Intermediate/Severe
PM1-9	**A**	A	A	A	**A**	**G**	-	1, 2, 4, 8, 9, 11, 15, 17, 20, 21, 24, 25, 30, 37, 40, 41, 44, 47, 50	9/10	Intermediate
PM1-14	**C**	A	**U**	**G**	**U**	**G**	G44A	14	7/10	Severe (4) No symptoms (3)
PM1-16	**C**	A	A	A	G	A	-	16	8/10	Severe
PM1-30	**A**	**C**	A	A	**A**	**G**	-	6, 20, 24, 27, 30, 48	10/10	Intermediate (8) No symptoms (2)
PM1-31	**C**	A	A	A	**A**	**G**	-	31	9/10	Severe
PM1-37	**A**	A	**C**	A	**A**	**G**	-	17, 37, 41, 47	20/22‖	Severe
PM1-40	U	A	A	A	**A**	**G**	-	11, 21, 40	10/10	Severe
PM1-43	**C**	A	**U**	A	**A**	A	U306A	43	3/10	Severe (1); Mild (2)
PM1-46	**A**	A	A	A	G	A	-	16, 46	9/10	Severe

*Mutagenized genome positions are numbered according the the PSTVd-DI reference isolate sequence. Nucleotides differing from the parental PSTVd-S23 sequence are indicated in bold.

†For each variant the code number of the plant(s) in which the variant was observed are indicated.

587 ‡The values given represent the number of plants showing PSTVd accumulation over the total
588 number of tomato plants inoculated with plasmids harboring a full-length infectious cDNA of
589 the relevant variant.
590 §Symptoms observed at 5 wpi. on the tomato plants inoculated with plasmids harboring a full-
591 length infectious cDNA of the relevant variant. When plants showing symptoms of different
592 severity were observed, the number of plants with each type of symptoms is indicated
593 between brackets.
594 ‖Sequence analysis of the progeny in four plants provided indications of C/A sequence
595 heterogeneity at position 52 (which differentiates PM1-37 from PM1-9) for three plants and
596 the presence only of PM1-9 for the fourth one.
597

FIGURE LEGENDS

Figure 1. Schematic representation of the predicted rod-like secondary structure of the PSTVd-S23 genome showing the two regions targeted for mutagenesis. Nucleotides mutagenized in the TL and pre-melting loop 1 (PM1) regions are underlined. The five domains of the PSTVd molecule are indicated, together with PM1.

Figure 2. Predicted secondary structures of the TL domain of recovered PSTVd-TL variants. The mutagenized positions and positions diverging from the parental PSTVd-S23 isolate are shown in bold and underlined, respectively. (A) classical folding; (B) alternative branched folding of the same region, shown for PSTVd-S23 and for variant TL4/11.

Figure 3. Accumulation of PSTVd-TL variants and PSTVd-S23 in 'Rutgers' tomato plants. PSTVd accumulation was detected by molecular hybridization with an *in vitro* transcribed PSTVd RNA probe. Detection was performed at 4 and 5 wpi. C-: negative control, plant inoculated with *Agrobacterium* carrying the empty pGreen vector; C+: positive control, PSTVd-infected plant; Buf.: extraction buffer spotted on the membrane.

Figure 4. **Predicted secondary structures of the recovered PM1 variants.** The mutagenized positions are shown in bold when identical to the parental S23 sequence and underlined and in bold when diverging from it. Additional mutations of some variants are also underlined and on bold. The predicted ΔG value calculated for the complete molecule is indicated.

Figure 1.

A

PSTVd-S23

```
  G      A    C
C GAACU AACU GUGGUUCC 21
  |||||  ||||  |||||||||
U CUUGG-UUGA-CGCCAAGG 341
  C
```

PSTVd-TL34/3

```
  G      A    C
A GAACU AACU GUGGUUCC 21
  |||||  ||||  |||||||||
A CUUGG-UUGA-CGCCAAGG 341
  A
```

PSTVd-TL34/7

```
  G      A    C
A GAACU AACU GUGGUUCC 21
  |||||  ||||  |||||||||
A CUUGG-UUGA-CGCCAAGG 341
  G
```

PSTVd-TL4/11

```
  GC     A    C
A  AACU AACU GUGGUUCC 21
   ||||  ||||  |||||||||
A  UUGG-UUGA-CGCCAAGG 341
  CA
```

B

PSTVd-S23

```
          C U
         A   C
         A   G
          A-U
          U-G
          C-G
          A-U
          A-U
          G-C
          G-C
         C   U
  U      U     U
G UGGUUCC     G GGUU 28
|  |||||||     |  ||||
A GCCAAGG      C-CCGA 341
  C            U U
```

PSTVd-TL4/11

```
          C U
         A   C
         A   G
          A-U
          U-G
          C-G
          A-U
          A-U
         C   C
         G   C
         A   U
  U      ACA   U
G UGGUU       G GGUU 28
|  |||||       |  ||||
A GCCAA        C-CCGA 341
  C      G     U
         G-U
```

Figure 2.

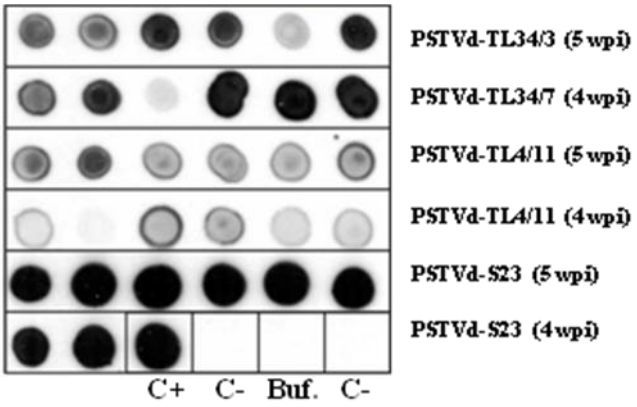
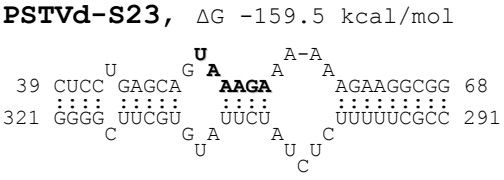
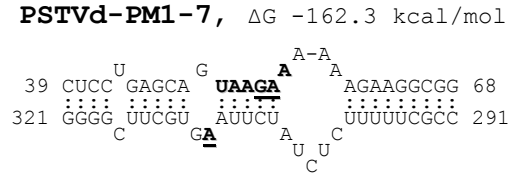


Figure 3.

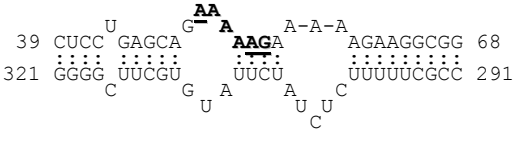
PSTVd-S23, ΔG -159.5 kcal/mol



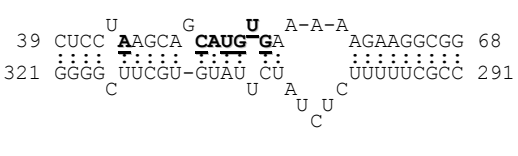
PSTVd-PM1-7, ΔG -162.3 kcal/mol



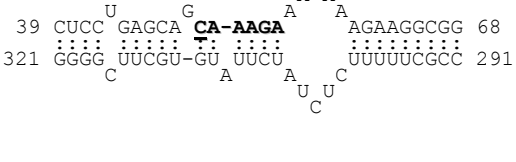
PSTVd-PM1-9, ΔG -158.4 kcal/mol



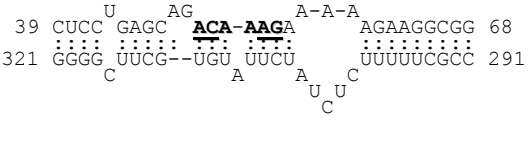
PSTVd-PM1-14, ΔG -160.6 kcal/mol



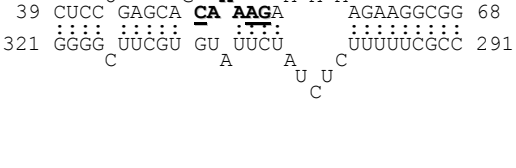
PSTVd-PM1-16, ΔG -160.5 kcal/mol



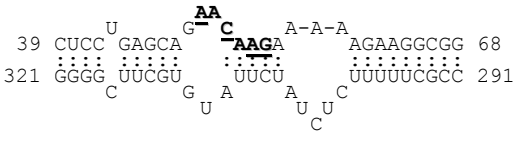
PSTVd-PM1-30, ΔG -158.5 kcal/mol



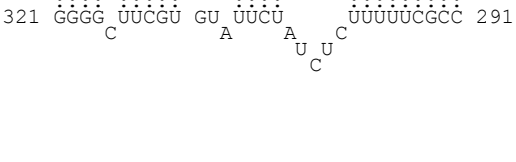
PSTVd-PM1-31, ΔG -161.3 kcal/mol



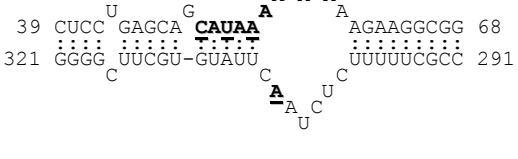
PSTVd-PM1-37, ΔG -158.4 kcal/mol



PSTVd-PM1-40, ΔG -159.4 kcal/mol



PSTVd-PM1-43, ΔG -160.5 kcal/mol



PSTVd-PM1-46, ΔG -159.5 kcal/mol

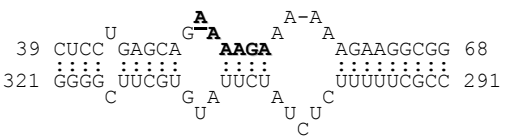


Figure 4.