



OPEN

## Role of an RNA pseudoknot involving the polyA tail in replication of *Pepino mosaic potexvirus* and related plant viruses

René C. L. Olsthoorn<sup>1✉</sup>, Carolyn A. Owen<sup>2</sup> & Ioannis C. Livieratos<sup>2</sup>

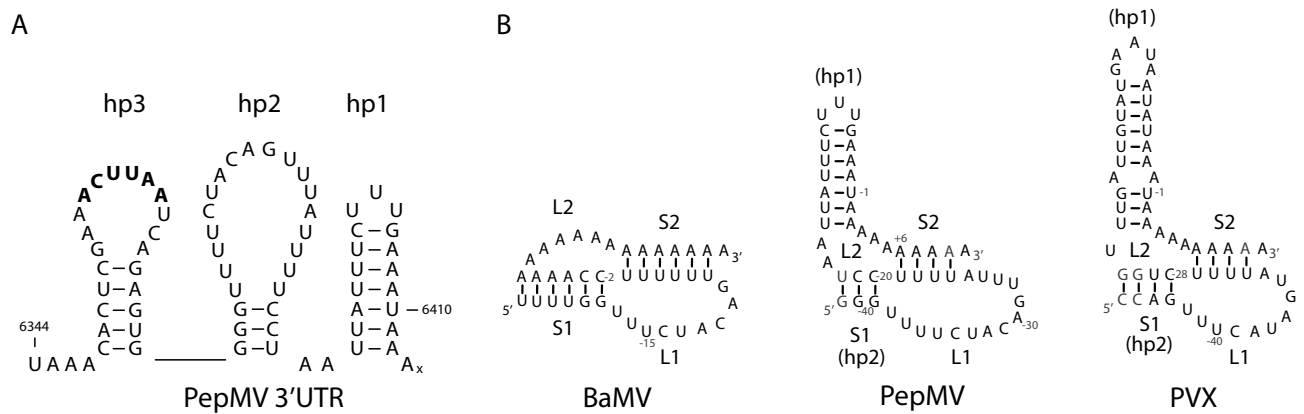
*Pepino mosaic virus* (PepMV) is a potexvirus of the family *Alphaflexiviridae* within the order of *Tymovirales* that threatens tomato production worldwide. PepMV possesses a positive-strand RNA genome with a 5'-methylguanosine cap and a 3'-polyA tail. Previously, using partially-purified viral RNA polymerase important secondary structures within the 3'-untranslated region (UTR) of PepMV RNA were identified. Here we show that an RNA pseudoknot can be formed in the 3'-UTR that includes part of the polyA tail. Using protoplasts, we demonstrate that the pseudoknot is required for replication of PepMV RNA. Mutational analysis and native gel electrophoresis further show that the pseudoknot is stabilized by UAU base triples, as is the human telomerase RNA pseudoknot. The presence of a pseudoknot in several other members of the *Alpha-* and *Betaflexiviridae* is supported by covariance analysis and native gel electrophoresis of other potexvirus, capillovirus and trichovirus RNAs. The ubiquitous presence of the pseudoknot in viruses of the *Betaflexiviridae*, suggests that the pseudoknot is a typical trait of the *Betaflexiviridae* that may have been adopted by many potexviruses during evolution.

Plant viruses belonging to the order *Tymovirales* are currently subdivided into five families: *Alpha-*, *Beta-*, *Gamma-* and *Deltaflexiviridae*, and *Tymoviridae*<sup>1</sup>. All members possess a single plus-stranded RNA genome of 6–9 kilobases (kb) that encodes an RNA-dependent RNA polymerase (RdRP), a coat protein (CP), and one or more proteins involved in cell-to-cell movement of the virus. RdRP is translated from the genomic RNA, CP is translated from a subgenomic (sg) RNA, and the movement proteins are expressed from a gene that overlaps with the RdRP gene or from additional sgRNAs. Members of the *Tymoviridae* are distinct in that they form icosahedral-shaped particles and possess a tRNA-like structure (TLS) at the 3'-end of the RNA, which is usually charged with valine. *Alpha-*, *Beta-*, *Gamma-* and *Deltaflexiviridae* are characterized by flexuous, filamentous-shaped particles that are 470–1000 nm in length and 12–13 nm in diameter. Their genomes lack a TLS but have a 3'-polyA tail, which for *Bamboo mosaic potexvirus* (BaMV) has been shown to fold into a pseudoknot structure. So far, a pseudoknot structure has not been identified in other potexviruses or any other member of the *Alpha-*, *Beta-*, *Gamma-* and *Deltaflexiviridae*.

*Pepino mosaic virus* (PepMV) is a mechanically-transmitted potexvirus (*Alphaflexiviridae*) with an approximately 6.4 kb genome with a 5'-methylguanosine cap and a 3'-polyA tail<sup>2</sup>. The genome contains five open reading frames, coding for an RdRP, three proteins involved in virus movement, and CP, flanked by 5'- and 3'- untranslated regions (UTRs) of ~86 and ~64 nucleotides (nts), respectively<sup>3</sup>. In vivo and in vitro experiments have shown that three 3'-co-terminal sgRNAs are produced that express CP and the other proteins<sup>4,5</sup>.

So far, secondary and tertiary structures have been identified in the 3'-UTRs of two other potexvirus genomes<sup>6–8</sup>. In the case of the type member *Potato virus X* (PVX) 3'-UTR, the RNA folds into three stem loops (hp1, hp2, hp3; numbered from the 3'-end<sup>8</sup>). hp1 is not required for minus-strand synthesis but hp2, which harbours a U-rich loop, is essential for plus- and minus-strand synthesis<sup>8,9</sup>. The BaMV 3'-UTR folds into a cloverleaf-like structure of four hairpins, followed by a pseudoknot that incorporates approximately 13 adenines of the polyA tail<sup>6,7</sup>. The PepMV 3'-UTR folds into three stem-loop structures (Fig. 1A). In vitro assays using partially-purified viral polymerase have shown that hp1 is dispensable for minus-strand synthesis. hp3, which harbours the potexvirus hexamer, and hp2 are required for minus-strand synthesis. These two hairpins

<sup>1</sup>Leiden Institute of Chemistry, University of Leiden, P.O. Box 9502, 2300 RA Leiden, The Netherlands. <sup>2</sup>Department of Sustainable Agriculture, Mediterranean Agronomic Institute of Chania, Alysio Agrokepion, 73100 Chania, Crete, Greece. ✉email: olsthoor@chem.leidenuniv.nl



**Figure 1.** Secondary structure of PepMV 3'-UTR and representative pseudoknot structures of potexvirus RNAs. **(A)** Secondary structure of the 3'-UTR of PepMV isolate SP13. The stop codon of the CP ORF is boxed. The potex hexamer is indicated in boldface. The polyA tail starts after nt 6410. **(B)** BaMV pseudoknot and putative PepMV SP13 and PVX pseudoknots. The number of A residues in the tail is arbitrarily chosen. Please note: in contrast to the BaMV pseudoknot model of Tsai et al.<sup>7</sup> we have positioned 6 As instead of 3 in L2 as the distance to bridge the minor groove of a 6-bp stem is  $\sim 30$  Å whereas one nt generally bridges  $\sim 6$  Å<sup>11</sup>, and in addition it may be assumed that a sixth A might interact with the first U in L1.

are very similar to their PVX counterparts, and accordingly, mutations in the U-rich loop of the PepMV hp2 were detrimental to minus-strand synthesis *in vitro*<sup>10</sup>.

Here we demonstrate that the U-rich motifs are involved in the formation of a pseudoknot that includes the polyA tail. Using protoplasts, we show that the pseudoknot is required for the replication of PepMV RNA. Mutational analysis and native gel electrophoresis show that the stabilization of the pseudoknot by UAU base triples is analogous to that of the human telomerase RNA pseudoknot. The presence of a telomerase-like pseudoknot in several other members of the *Flexiviridae* is supported by covariance analysis and by native gel electrophoresis of other potexvirus, capillovirus and trichovirus RNAs.

## Results

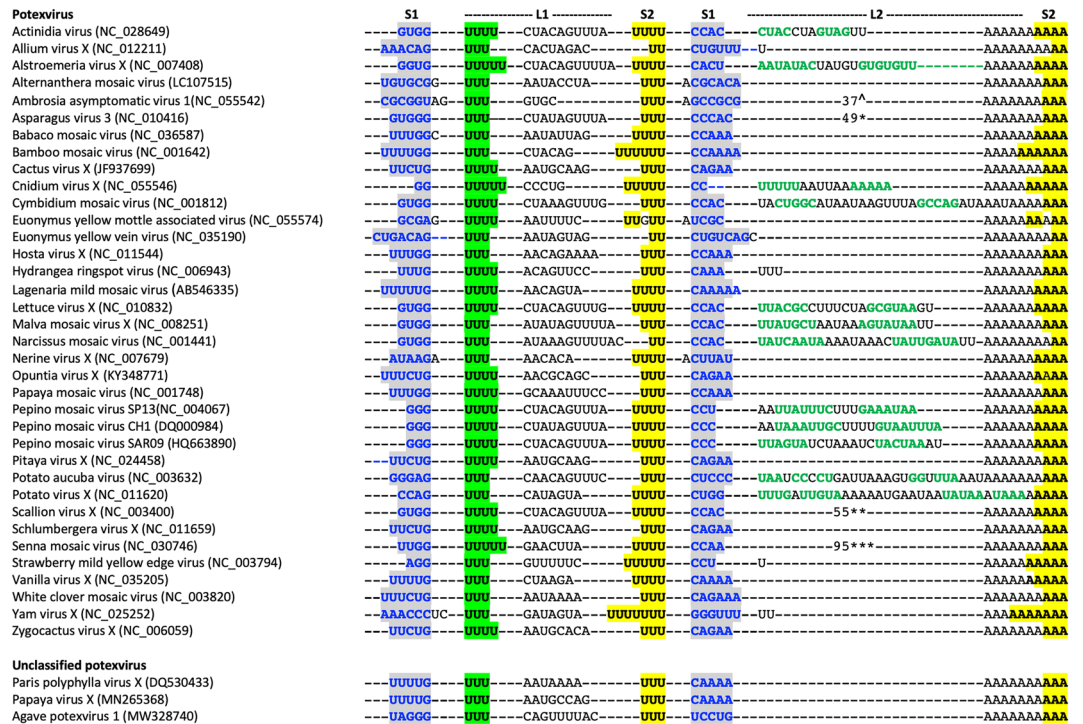
### Comparison of the 3'-UTRs of potexviruses shows potential pseudoknot formation involving the polyA tail.

Previously, the 3'-UTR of BaMV RNA was shown to terminate in a pseudoknot involving approximately 13 adenosines of the polyA tail (Fig. 1B<sup>7</sup>) but an analogous structure had not been detected nor proposed for any related potexviruses. Alignment of potexvirus RNA sequences, however, shows that a stem equivalent to stem S1 of the BaMV pseudoknot can be formed in the majority of potexviruses, including three unclassified potexviruses (Fig. 2, blue characters); while equivalent sequences could not be detected in a further 13 potexvirus species. The existence of S1 is further supported by the presence of covariations; almost every possible base-pair can be found in S1, indicating that base-pairing potential of bases at these positions in the genome is preserved among these viruses. An exception may be the strongly conserved GC pair closest to loop L1; in only a few potexviruses is this pair CG, AU, GU, or UG. In approximately half of these potexvirus RNAs, adenosines from the polyA tail are involved in the formation of S1, while in the remainder the polyA tail is not involved in S1 and is preceded by 14–95 nts with the ability to form an AU rich stem-loop structure (indicated in green font).

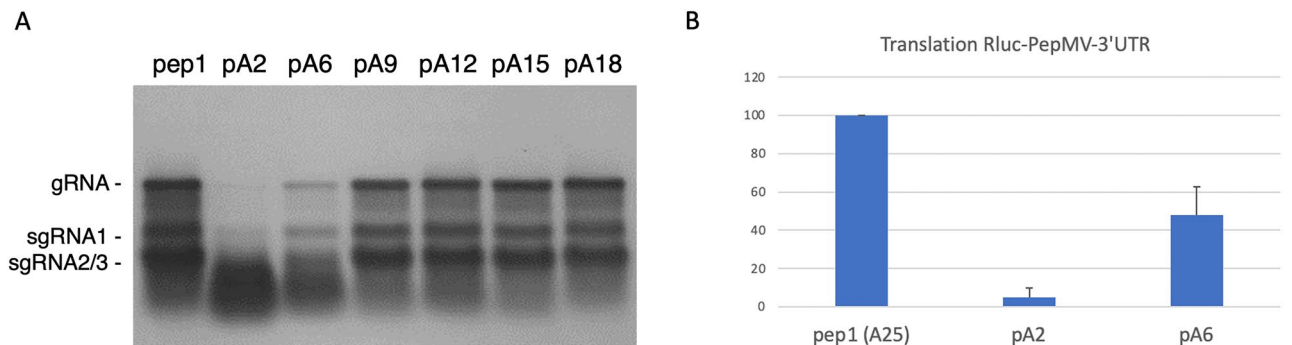
By analogy with the BaMV pseudoknot, all sequences contain U-rich motifs that flank the nucleotides of stem 1 (highlighted in green and yellow). For BaMV it was shown that the 3'-proximal Us (highlighted in yellow) can base-pair with 6 As of the polyA tail (highlighted in yellow), thereby forming stem S2. In the potexvirus RNAs shown here, 2–7 Us are available for base-pairing with As from the polyA tail. The conserved 5'-proximal U-rich motif (highlighted in green) can in principle also base-pair with As from the polyA tail but this would lead to an unusual and rare pseudoknot structure wherein the polyA tail has to span both S1 and S2. Since we found that the 5'-proximal Us are involved in base triples (see below) we did not consider this alternative pseudoknot conformation.

Potexvirus pseudoknots can thus be grouped into those that have additional sequences capable of forming a stem-loop structure within the loop L2 of the pseudoknot (e.g. PepMV and PVX), and those for which L2 is composed of As from the polyA tail (e.g. BaMV). A few viruses have one (*Allium virus X* and *Strawberry mild yellow edge virus*), two (*Yam virus X*) or three Us (*Hydrangea ringspot virus*) that separate S1 from the polyA tail. While the role of this additional structure remains unknown and may be host-related, its presence is probably not required for the synthesis of minus- or plus-strand RNA, as was reported for PVX<sup>8</sup>. Also, replication of PepMV RNA *in vitro* and in protoplasts is not affected by removal of this hairpin<sup>10</sup>.

**Effect of polyA-length on replication *in vivo*.** The length of the polyA tail is obviously crucial for the formation of the pseudoknot and indeed BaMV transcripts with 10 or less As are not infectious as they probably do not allow formation of the pseudoknot<sup>7</sup>. To investigate the length of the polyA tail required for replication of PepMV, a full-length PepMV SP13 cDNA clone was used as template for PCR mutagenesis and *in vitro* tran-



**Figure 2.** Alignment of 3'-ends of potexviruses capable of forming a pseudoknot structure. Bases in blue font form stem S1, those highlighted in yellow form stem S2. The 5'-proximal U-rich motif is highlighted in green. Putative hairpins formed by bases in L2 are shown in green font. \*AGCAGACUAUCAUUAUUCACUCUCU UUGAGCGUAAUUAAGUACGUGUGUU insert in L2 of *Asparagus virus 3*. \*\*AUCAGACUCCUAUCCUA CUAGCUUUAUCCGCAUGUAUGAAUGUAAGUUUGUUUU insert in L2 of *Scallion virus X*. \*\*\*UUCUUG CCACCGCCAGAGUGAGAGUCUAGUUUAGUCAGCCCUGUUUUUCGCACUUUUGUUGGGGCUAU UGAGUUUUCAAAAGUGCUGUCUAGCU insert in L2 of *Senna mosaic virus*.

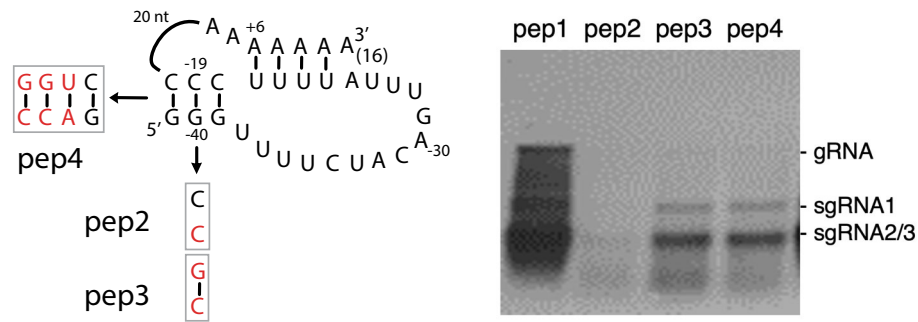


**Figure 3.** The effect of polyA tail length on the replication of PepMV. (A) Northern blot showing the accumulation of PepMV gRNA and sgRNAs in *N. benthamiana* protoplasts. pep1: wt PepMV with A25, A2, A6, A9, A12, A15, A18. The migration of gRNA and sgRNAs is indicated. We note that sgRNA2 and 3 comigrate on these gels. (B) Effect of polyA tail length on translation efficiency of the PepMV-3'UTR in protoplasts. Error bars indicate standard deviation of 2 experiments.

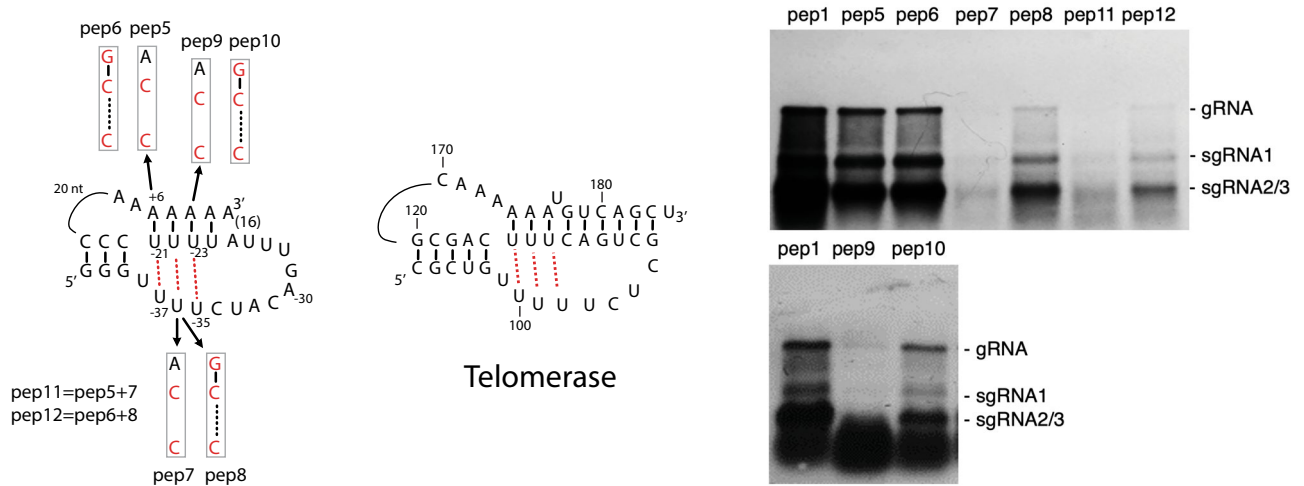
scription. In the final capped mRNAs U<sub>-18</sub> has been substituted by a C, a natural variation found in several other PepMV isolates (Fig. 2). This substitution is thought to stabilize stem 1 of the pseudoknot<sup>10</sup>.

Transfection of *Nicotiana benthamiana* mesophyll protoplasts with PepMV RNA with a tail of only 2 As (pA2) led to barely detectable accumulation of gRNA and sgRNAs (Fig. 3A). However, the presence of 6 As (pA6) resulted in significantly enhanced replication while with 9 As (pA9) the replication levels were close to those of the reference construct incorporating 25 As (pep1). The presence of 12, 15, or 18 As did not result in a detectable further increase in replication. The construct with 25 As (pep1) will be referred to as wild type (wt) henceforth.

As the polyA tail is also necessary for translation, an essential early process in PepMV infection, it is likely that the number of As will also affect translation. In an attempt to differentiate the effects of polyA tail-length on replication and translation, templates were made in which a *Renilla* luciferase gene was fused to the PepMV 3'



**Figure 4.** Role of S1 length and stability in replication. For the sake of clarity hp1 is not depicted. Northern blot showing the accumulation of PepMV gRNA and sgRNAs in *N. benthamiana* protoplasts.



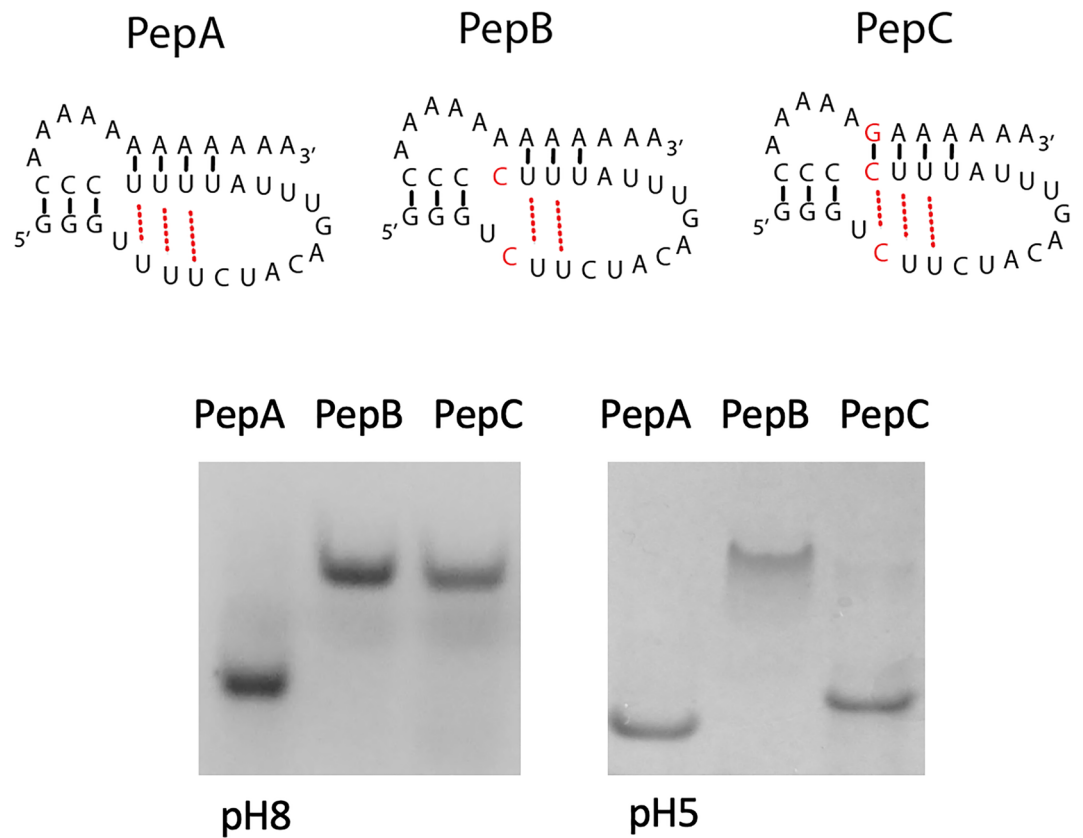
**Figure 5.** Putative UAU triples in PepMV by analogy to the human telomerase pseudoknot. Northern blots showing the accumulation of PepMV gRNA and sgRNAs in *N. benthamiana* protoplasts after transfection with the indicated constructs.

UTR that incorporated a polyA tail of 2, 6, or 25 As. Capped RNA transcripts synthesized from these templates were transfected into *N. benthamiana* protoplasts and luciferase activity as a measure of protein translation was determined 16 h post-transfection. The presence of 2 As resulted in very low levels of translation (~5%), whereas translation of the transcript incorporating 6 As was only twofold less than the wt RNA with 25 As (Fig. 3B). This result suggests that the low replication observed with a polyA tail of 6 nts is mainly attributable to a decrease in minus-strand RNA synthesis, rather than to diminished translation.

**Stem S1 requirements.** Previously, we demonstrated that base pairing in hp1 was important for the replication of PepMV RNA in vitro. Here we investigated whether the integrity of stem S1 (*i.e.* the stem of hp2) is important for replication in vivo. Disruption of the middle GC bp to a CC mismatch (pep2) greatly reduced replication. (Fig. 4). Restoring S1 by converting the CC mismatch to a CG bp (pep3) improved replication to some extent but the level was substantially lower than for the wt. Replacing S1 with the corresponding stem of PVX (pep4) also resulted in similarly reduced replication. This is potentially due to the altered stability of S1: pep3 – 5.80 kcal/mol, pep4 – 7.50 kcal/mol, wt – 6.60 kcal/mol (calculated by Mfold<sup>12</sup>), although a loss or change of putative tertiary interactions could also contribute to a lower accumulation. Note that a CG pair as in pep3 is rarely found as the penultimate bp in stem 1 of potexviruses (Fig. 2).

**Role of putative UAU base triples in replication.** Pseudoknots are stabilized by various factors, including interactions between bases in the loops and base pairs in the stems<sup>13</sup>. For example, UAU base triples formed between U residues in the loop and AU bps in the stem have been shown to be essential for the structure and stability of the human telomerase pseudoknot (Fig. 5<sup>14</sup>) and of other noncoding RNAs<sup>15</sup>. In these studies, UAU base triples could be functionally replaced with isosteric CGC triples. To investigate the presence of UAU base triples in PepMV we disrupted three potential UAU triples identified from homology with the human telomerase pseudoknot, by replacing Us with Cs (pep5, pep7, pep9). Replication of pep5 was moderately reduced, while replication of pep7 and pep9 was severely reduced (Fig. 5). By introducing an A-to-G change in the polyA tail of these mutants we attempted to restore base-pairing in stem S2 of the pseudoknot and generate an isomorphous CGC base triple (Fig. 5; pep6, pep8, pep10). This introduction did not result in increased



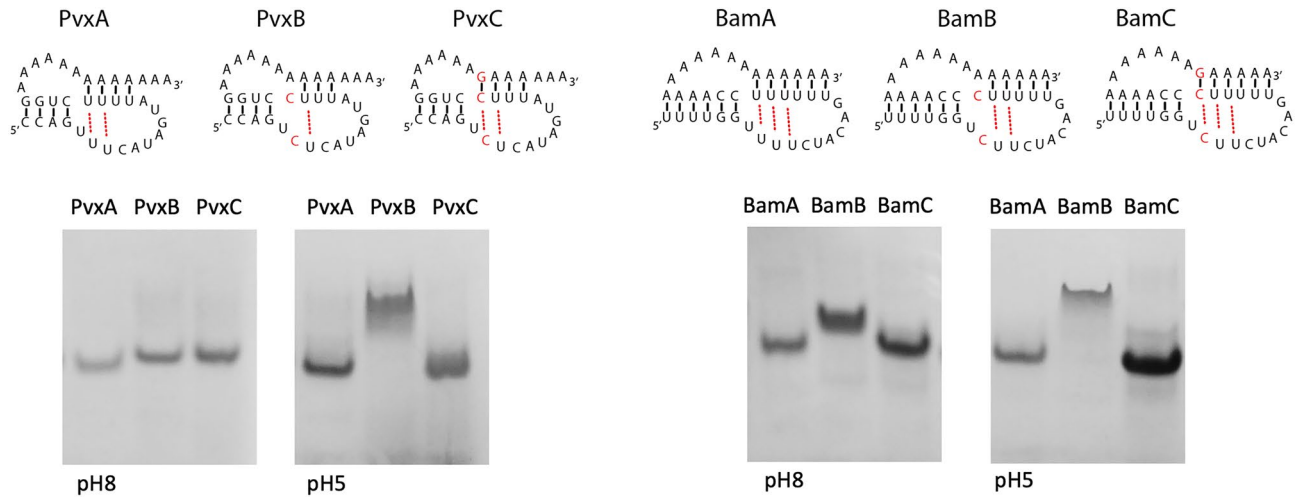


**Figure 6.** Native gel electrophoresis at pH 8 and 5 of PepMV pseudoknot RNAs with wild type (PepA) or mutant base triples (PepB, PepC). RNAs were visualized by Stains-All.

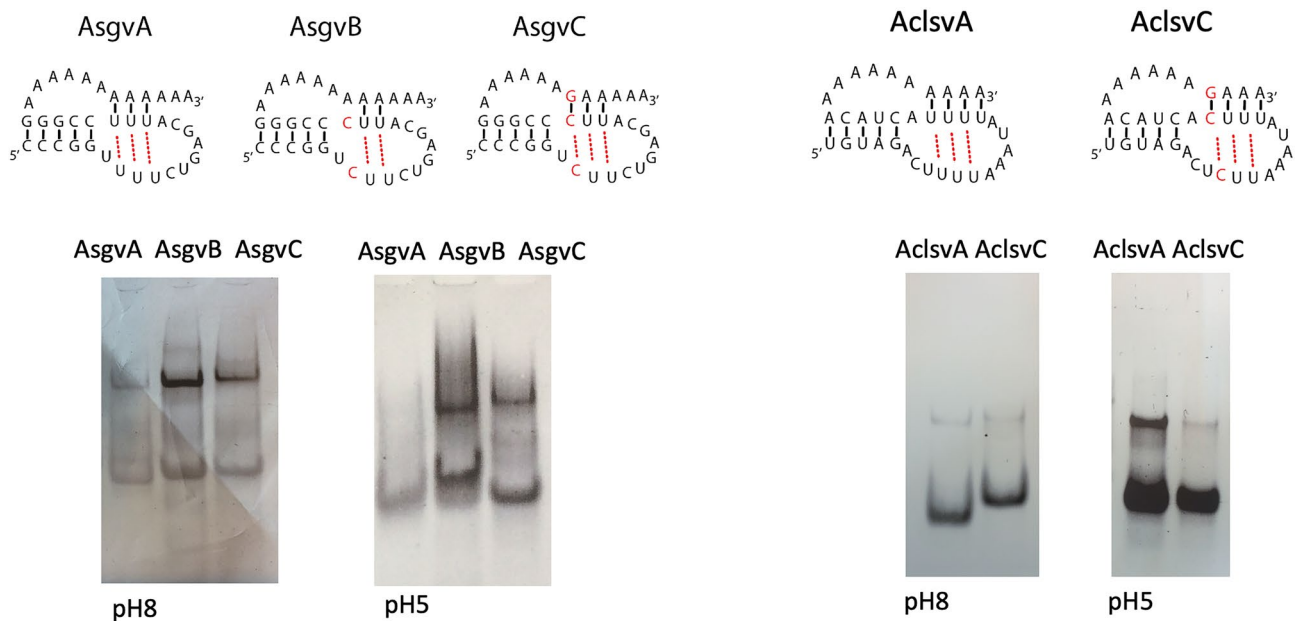
accumulation of PepMV RNAs for pep6 compared to pep5, but did lead to a substantial increase in accumulation of PepMV RNAs for pep8 and pep10 relative to their respective CAC mutants pep7 and pep9 (Fig. 5). The simultaneous mutation of the triples  $A_{+6}U_{-21}U_{-37}$  and  $A_{+7}U_{-22}U_{-36}$  yielded results similar to mutation of  $A_{+7}U_{-22}U_{-36}$  alone (Fig. 5, compare pep11 and pep7) while introduction of CGC triples at these positions again partially restored replication (Fig. 5, pep12). These data strongly suggest that UAU triples are formed between  $A_{+6}U_{-21}U_{-37}$ ,  $A_{+7}U_{-22}U_{-36}$  and  $A_{+8}U_{-23}U_{-35}$ , in a structure analogous to the human telomerase pseudoknot, but that the  $A_{+6}U_{-21}U_{-37}$  triple is quite tolerant of mutation.

**Native gel electrophoresis of model potexvirus pseudoknots.** To study the importance of base triples for the structure of PepMV and other potexvirus pseudoknots, we used short RNA oligonucleotides and native gel electrophoresis. In PepA (36 nts) which corresponds to the PepMV pseudoknot from which hp1 has been removed, potentially three UAU base triples can be formed, whereas in PepB two U-to-C changes disrupt one base triple, and in PepC, in addition to these two U-to-C changes, an A-to-G change in the polyA tail is introduced to create a CGC base triple. Figure 6 shows that PepA migrated faster than both PepB and PepC, indicating a more compact structure. Interestingly, at pH5 when cytosines become protonated, the migration of PepC equaled that of PepA, while PepB still lagged behind. Since the only difference between PepB and PepC is the substitution of one G for an A in the tail, the most likely explanation for their difference must be the formation of a CGC triple. 1D proton NMR spectroscopy of PepA also supported the existence of a pseudoknot structure for this RNA (Supplementary Fig. S1). We note that CGC triples are naturally found in a variety of RNAs and can substitute for UAU triples but that in vitro protonation of one of the cytosines is generally required for their formation.

Similar results were obtained with PvxA, using a model oligonucleotide of 34 nts corresponding to the 3'-end of PVX RNA but lacking the terminal hairpin that is not required for minus-strand synthesis<sup>8</sup>. At pH 8 PvxA migrated faster than PvxB which has two U to C changes that disrupt the UAU triple, and PvxC with the CGC triple (Fig. 7). Again, at low pH PvxC co-migrated with PvxA while PvxB did not. The BaMV pseudoknot (BamA) migrated faster than the double mutant (BamB) demonstrating that base triples are also present in BaMV RNA. Interestingly, BamC with the CGC triple migrated as fast as BamA at both high and low pH, suggesting that in the context of the BaMV sequence the CGC triple is stable at pH 8. In conclusion, PVX and BaMV RNAs also adopt a pseudoknot structure that is stabilized by UAU base triples.



**Figure 7.** Native gel electrophoresis at pH 8 and 5 of PVX and BaMV pseudoknot RNAs. RNAs were visualized by Stains-All.



**Figure 8.** Native gel electrophoresis at pH 8 and 5 of *Apple stem grooving capillovirus* (AsgvA, B, C) and *Apple chlorotic leaf spot trichovirus* (AclsvA and C) pseudoknot RNAs. RNAs were visualized by EtBr staining.

**Pseudoknot conservation in other Alpha- and Betaflexiviridae.** In order to establish whether the 3'-UTRs of other *Alpha-* and *Betaflexiviridae* could adopt a similar pseudoknot, an alignment was made of all available sequences from the *Alphaviridae*. This alignment indicated that pseudoknots may exist in the genus *Mandarivirus*, but not in *Allexi-*, *Botrex-*, *Lola-*, *Platypu-* and *Sclerodarnaviruses* (see full alignment supplementary Fig. S2). Among the *Betaflexiviridae*, the pseudoknot was found within the subfamily *Quinvirinae* in the genera *Carlavirus*, *Foveavirus*, and *Robigovirus*, as well as in some unclassified species. Within the subfamily *Trivirinae* similar pseudoknots could be found in all genera except for *Vitivirus*. In all but 4 viruses the pseudoknot stems S1 and S2 consists of at least three base pairs and the majority have the potential to form three UAU triples. In contrast to those of the potexviruses a large number of these pseudoknots possess a single C or A residue between stems S1 and S2. This would not necessarily prevent pseudoknot formation as the existence of unpaired bases at the junction between two coaxially stacked stems of a pseudoknot is not uncommon<sup>16</sup>. A few sequences, mostly from trichoviruses, feature unpaired bases separating the S1 and the U-stretch involved in UAU base triples.

We tested whether pseudoknot formation is possible for the capillovirus *Apple stem grooving virus* (AsgvA), and the trichovirus *Apple chlorotic leaf spot virus* (AclsvA) which has an A at the junction between the stems and 2 nts upstream of the U-stretch (Fig. 8). Both wt RNAs, AsgvA and AclsvA, migrated faster through high pH gels

than mutants in which the UAU triple was replaced by a CGC (AsgvC, AclsvC) or a CAC mismatch (AsgvB). At low pH, migration of wt and the CGC mutants was identical, indicating that the protonated CGC triple restored the pseudoknot, whereas the CAC mutant (AsgvB) migrated even more slowly. 1D proton NMR spectroscopy of AsgvA and AsgvB samples also showed clear differences in the area of UA base pairs, indicative of pseudoknot formation in AsgvA (Supplementary Fig. S3). These results show that a pseudoknot can be formed in capillo- and trichoviruses and that intervening nucleotides between stems do not interfere with pseudoknot formation.

## Discussion

To date, the only member of the *Potexvirus* genus shown to possess a 3'-terminal pseudoknot was BaMV. However, the data in this work strongly suggest the presence of a pseudoknot structure involving the polyA tail in the RNA of PepMV, and also in the majority of the potexviruses. The PepMV pseudoknot is very likely essential for replication of the RNA as translation was less sensitive to the length of the polyA tail. With a 6 A-tail translation was only twofold lower than wt (25 As), whereas replication was strongly reduced. On the other hand, 9 As were sufficient to obtain wt levels of RNA replication in vivo. This contrasts with the results of previous in vitro experiments using partially purified PepMV replicase that showed no detectable minus strand synthesis with polyA tails of 10 or 15 As<sup>5</sup>. This discrepancy may be due to the addition of As in vivo by a cellular polymerase, acting in competition with the degradation of templates with short polyA tails. Folding of the 3'-terminus into a pseudoknot with 6, 10 or 15 As may protect against this degradation in vivo and allow further extension of the polyA tail to produce a replicable template.

For BaMV, no replication in vivo was detected with a polyA tail of 10 A or less. Since in BaMV 4 As are necessary for the formation of stem S1 of the pseudoknot, the remaining 6 As are apparently not sufficient to make up stem S2 and loop L2. The minimum length of the polyA tail for replication of another well-studied potexvirus, PVX has not been definitively determined: 8 As were reported to yield very low replication but the transcript also incorporated 4 additional nts derived from a restriction site<sup>17,18</sup>. Pillai-Nair et al.<sup>8</sup> achieved successful infection of protoplasts using PVX transcripts with 17 As that also contained the 4 additional nts. In any case, as 3 As form part of hp1, the remaining 14 As are sufficient to form the proposed pseudoknot for PVX.

Mutational analysis and native gel electrophoresis also indicated that the PepMV pseudoknot is stabilized by a number of UAU base triples, as their disruption reduced replication whereas their substitution with isosteric CGC base triples preserved replication. We observed that not all triples are equally important. The triple closest to stem S1 can be disrupted by CAC (pep5) or replaced by CGC (pep6) without much effect on replication. On native gels, among PepMV RNA oligonucleotides lacking hp1, mutation of the triple closest to S1 was sufficient to change the migration and structure of the pseudoknot, and alter the NMR spectrum. Whether hp1 exerts a stabilizing effect on the pseudoknot and possibly mitigates the effect of destabilizing mutations remains to be investigated.

CGC triples did not, however, fully restore replication. One reason could be that CGC triples require protonation of one of the cytosines which at neutral pH occurs slowly but can be accelerated by neighbouring UAU triples<sup>19</sup>. Other possible reasons for the lower levels of replication are: (i) the PepMV triples are in a different register and so were here inadvertently introduced at the wrong position, thereby forcing the structure into a conformation less favorable for replication. At present, we cannot rule out this possibility, but the fact that constructs with CGC triples replicated better than their corresponding CAC mismatches at least indicates that the PepMV pseudoknot is stabilized by UAU triples. Elucidation of the 3D structure of the PepMV pseudoknot will help to determine the exact nature of these triples and also the extent of the homology with the human telomerase pseudoknot, or with other pseudoknots stabilized by UAU triples<sup>20</sup>. (ii) The interruption of the polyA tail by one or more G may adversely affect translation leading to decrease in viral protein levels and, indirectly, to less synthesis of gRNA and sgRNAs. Currently, no data exist to support this hypothesis. (iv) The CGC triple may interfere with the formation of an alternate hairpin structure which in the wt is stabilized by UU base pairs, but in the mutants these are replaced by CC or CU pairs. This hairpin structure may be the binding site for a host factor such as that shown by Sriskanda et al. to bind to the U-rich sequence in L1 of PVX<sup>9</sup>, although it cannot be excluded that this factor recognizes the UAU triples, as mutation of these Us would also disrupt this interaction.

Based on the structural resemblance to the PepMV pseudoknot, and data from native gel electrophoresis, we also propose a 3'-terminal pseudoknot involving the polyA tail for PVX. Although previous structure-probing data on the PVX 3'-UTR did not show clear evidence of the U-rich motifs being involved in base pairing, Pillai-Nair et al.<sup>8</sup> found U<sub>-40</sub>, U<sub>-41</sub> and U<sub>-42</sub>, that we predict to be involved in the formation of the UAU base triples, to be largely protected from modification by CMCT. The 4 Us involved in the proposed AU bps in stem S2, however, were more accessible to CMCT but still less so than other Us in hairpin loops in the 3'-UTR of PVX RNA. Interestingly, U<sub>-15</sub> to U<sub>-17</sub> in loop L1 of the BaMV pseudoknot are protected from RNase A cleavage but are a substrate for the dsRNA specific RNase V1<sup>9</sup> which supports the presence of UAU triples in the BaMV pseudoknot as well. Other data that support the presence of a pseudoknot in PVX come from a study using BaMV-PVX hybrids<sup>21</sup> showing that the PVX 3'-UTR can partially substitute for the potex hexamer and pseudoknot domains of BaMV in in vitro minus-strand synthesis. Future research should demonstrate the requirement of the pseudoknot for the replication of PVX.

The pseudoknot can be found in 34 classified and three unclassified potexviruses but is absent in 13 classified potexviruses. These 13 viruses cluster together in five subclades based on similarities in their replicase and coat protein genes<sup>22</sup> which we have tentatively assigned here to group 1 *Plantago asiatica mosaic virus*, *Tulip virus X* and *Cassava common mosaic virus* [CasCMV]; group 2 *Lily virus X*, *Phaius virus X* and *Mint virus X*; group 3 (*Clover yellow mosaic virus* [CIYMV] and *Tamus red mosaic virus* [TaRMV]); group 4 (*Foxtail mosaic virus* [FxmV] and *Turtle grass virus X* [TGVX]); group 5 (*Cassava virus X*, *Cassava mosaic virus X* [CasMV], and *Cassava Colombian symptomless virus*. Careful inspection of their 3'-ends shows that group 1 shares a stem-loop

structure with certain marafiviruses of the *Tymoviridae* (Fig. S4). The 3'-ends of group twofold into a so-called pseudotri-loop hairpin which functions as a recognition element for *Brome mosaic virus* replicase<sup>23</sup> but is also found at the 3'-ends of some allexiviruses and the botrexvirus *Botrytis virus X* (Fig. S4). Presumably, during evolution these potexviruses exchanged the 3' pseudoknot with the 3' stem-loop of plant viruses from a nearby family. Such events have also been documented for other plant viruses<sup>24,25</sup> and avian caliciviruses<sup>26</sup>.

CIYMV and TaRMV (Group 3) both form a stem-loop structure with a similar loop sequence that is not found in other viruses (Fig. S4). Cassava viruses (group 5) also form a stem-loop structure with loop sequences that, apart from a CAG sequence, have little in common. Interestingly, the loop sequence of CasMV (ACAGUU UA) is very similar to loop L1 of the pseudoknot of other potexviruses like PepMV, *Actinidia virus*, *Alstroemeria virus X*, etc. Group 4 viruses FxMV and TGVX form a hairpin of 6–7 bp with a loop sequence AAUGCACA that shows strong homology with the loop sequence of the pseudoknot of *Zygocactus virus X*, *Cactus virus X*, *Schlumbergera virus X*, *Pitaya virus* (Fig. 2). This is remarkable as FxMV and TGVX are more closely related to BaMV, which has the CUACAG motif that is shared by potexviruses like PepMV. The exact function of this motif is not known, but the fact that satellite RNAs of BaMV<sup>27</sup> which are replicated by the same RdRP do not possess this motif suggest that it is bound by a host factor rather than by the RdRP itself. A potential host factor may be NbHsp90 that was shown to bind to the BaMV pseudoknot but not to the 3' end of its satellite RNA that lacks the pseudoknot<sup>28</sup>.

This type of pseudoknot involving the polyA tail can be found in one other genus, *Mandarivirus*, of the *Alphaflexiviridae*, but not in the genera *Allexivirus*, *Botrexvirus*, *Lolavirus*, *Platypuvirus*, and *Sclerodarnavirus* of the *Alphaflexiviridae*. The only allexivirus sequence that was found to harbour a pseudoknot is the unclassified allexivirus *Garlic yellow virus*, but this virus is closely related to *Allium carlavirus A*, an unclassified carlavirus. In the related *Betaflexiviridae* however, the pseudoknot is widespread: it is found in 11 genera, and is only not present in the genus *Vitivirus*. The 3'-ends of vitivirus RNAs can be folded into a hairpin that in many cases is stabilized by base pairs involving the polyA tail (Fig. S4) similar to allexiviruses. Thus the pseudoknot seems to be a typical trait of the *Betaflexiviridae* that was somehow adopted by many potexviruses during evolution. It will be of considerable interest to find out whether the pseudoknot also plays an important role in the replication of viruses belonging to the *Betaflexiviridae*.

## Methods

**Synthesis of RNA templates for protoplast inoculation.** DNA fragments representing differently-sized sequences of the PepMV plus-strand genome were amplified from the pTOPO-T7 PepXL6 DNA template. All PCRs were carried out using LA Taq DNA polymerase (TAKARA), the template DNA was restricted with *DpnI* prior to purification of the generated products by agarose gel electrophoresis and used as templates for in vitro transcription using the mMESSAGE mMACHINE T7 RNA kit (Ambion). The 5'-methylguanosine capped RNA transcripts produced were purified as described in the Ambion manual, and assayed by spectrophotometer before use in downstream experiments.

**PepMV replication assays.** Full length PepMV cDNA templates were amplified using a T7pep forward oligonucleotide primer and one of a set of 3'-co-terminal reverse oligonucleotide primers (Supplementary Table S1), purified and transcribed. Five µg of each transcript being used to inoculate  $5 \times 10^5$  *N. benthamiana* mesophyll protoplasts in the presence of PEG 4000, before incubation for 24 h at 25 °C under constant light. Protoplasts were harvested and total RNA was isolated using Trizol reagent (Invitrogen) according to the manufacturer's instructions. For subsequent northern blot analysis, 5 µg of each total RNA was resolved, transferred to charged nylon membranes and probed with a Dig-labelled (-) single stranded RNA probe corresponding to the PepMV CP gene coding sequence.

**Translation assays.** The RLuc-PepMV-3UTR plasmid was obtained by insertion of a DNA fragment covering the PepMV 3'UTR (Baseclear, Leiden, The Netherlands) into a *Renilla* luciferase reporter plasmid previously described<sup>29</sup>. Templates for transcription were obtained by PCR using the forward primer SP6FLU and reverse primer (pep1, pA2, or pA6—Supplementary Table S1). Transcription reactions were carried out as described above, but with the kit enzyme mix substituted with a 7:2:1 mixture of SP6 polymerase, recombinant RNase inhibitor, and inorganic pyrophosphatase (all New England Biolabs).  $5 \times 10^5$  *N. benthamiana* protoplasts were transfected with 3 µg of RNA and after 16 h incubation at 25 °C under constant light, were freeze-dried and sent by airmail for analysis at Leiden University. Upon arrival material was resuspended in 100 µl Tris (10 mM pH 8) and luciferase activity in 50 µl samples was measured using a GloMax multi system (Promega).

**Native PAGE.** RNA oligonucleotides were purchased from Merck (Sigma-Aldrich) at 0.05 nmole scale/desalted. 100–200 pmol of each RNA were loaded onto polyacrylamide gels containing 12% or 16% acrylamide:bisacrylamide (19:1), Tris (40 mM), acetate (20 mM), EDTA (1 mM) pH 8.3, with 1.5 mM MgAc<sub>2</sub>. Gels were run in TAEM buffer at ~85 V and 12 mA for ~4 h in a cold room. For acidic PAGE the pH was adjusted by the addition of acetic acid to all buffers. RNA was visualized using EtBr or Stains-All (Sigma Aldrich).

**NMR spectroscopy.** <sup>1</sup>H NMR experiments were recorded at 278° Kelvin on a Bruker 600 MHz spectrometer equipped with a cryoprobe using Watergate suppression. RNA oligonucleotides were dissolved in 10 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> buffer (pH 6.7), 0.5 mM MgSO<sub>4</sub> and 10% D<sub>2</sub>O to a final concentration of ~0.25 mM.



## Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Received: 17 May 2022; Accepted: 27 June 2022

Published online: 07 July 2022

## References

1. Virus Taxonomy: 2020 Release. International Committee on Taxonomy of Viruses (ICTV). March 2021 (accessed 15 May 2021).
2. Hanssen, I. M. & Thomma, B. P. H. J. *Pepino mosaic virus*: A successful pathogen that rapidly evolved from emerging to endemic in tomato crops. *Mol. Plant Pathol.* **11**, 179–189 (2010).
3. Aguilar, J. M., Hernandez-Gallardo, M. D., Cenis, J. L., Lacasa, A. & Aranda, M. A. Complete sequence of the *Pepino mosaic virus* RNA genome. *Arch. Virol.* **147**, 2009–2015 (2002).
4. Sempere, R. N., Gómez, P., Truniger, V. & Aranda, M. A. Development of expression vectors based on *Pepino mosaic virus*. *Plant Meth.* **7**, 6 (2011).
5. Osman, T. A., Olsthoorn, R. C. L. & Livieratos, I. C. In vitro template-dependent synthesis of *Pepino mosaic virus* positive- and negative-strand RNA by its RNA-dependent RNA polymerase. *Virus Res.* **167**, 267–272. <https://doi.org/10.1016/j.virusres.2012.05.008> (2012).
6. Cheng, J. H., Peng, C. W., Hsu, Y. H. & Tsai, C. H. The synthesis of minus-strand RNA of bamboo mosaic potexvirus initiates from multiple sites within the poly(A) tail. *J. Virol.* **76**, 6114–6120 (2002).
7. Tsai, C. H. *et al.* Sufficient length of a poly(A) tail for the formation of a potential pseudoknot is required for efficient replication of bamboo mosaic potexvirus RNA. *J. Virol.* **73**, 2703–2709 (1999).
8. Pillai-Nair, N., Kim, K. H. & Hemenway, C. Cis-acting regulatory elements in the potato virus X 3'-non-translated region differentially affect minus-strand and plus-strand RNA accumulation. *J. Mol. Biol.* **326**, 701–720 (2003).
9. Srisankanda, V. S., Pruss, G., Ge, X. & Vance, V. B. An eight-nucleotide sequence in the Potato virus X 3' untranslated region is required for both host protein binding and viral multiplication. *J. Virol.* **70**, 5266–5271 (1996).
10. Osman, T. A., Olsthoorn, R. C. L. & Livieratos, I. C. Role of the *Pepino mosaic virus* 3'-untranslated region elements in negative-strand RNA synthesis in vitro. *Virus Res.* **190**, 110–117. <https://doi.org/10.1016/j.virusres.2014.06.018> (2014).
11. Pleij, C. W., Rietveld, K. & Bosch, L. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13**, 1717–1731. <https://doi.org/10.1093/nar/13.5.1717> (1985).
12. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
13. Gulyaev, A. P., Olsthoorn, R. C. L., Pleij, C. W. A. & Westhof, E. RNA Structure: Pseudoknots. In *Encyclopedia of Life Sciences* (Wiley, 2012).
14. Theimer, C. A., Blois, C. A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell.* **4**, 671–682. <https://doi.org/10.1016/j.molcel.2005.01.017> (2005).
15. Mitton-Fry, R. M., DeGregorio, S. J., Wang, J., Steitz, T. A. & Steitz, J. A. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **330**, 1244–1247. <https://doi.org/10.1126/science.1195858> (2010).
16. van Batenburg, F. H., Gulyaev, A. P., Pleij, C. W. A., Ng, J. & Oliehoek, J. PseudoBase: A database with RNA pseudoknots. *Nucleic Acids Res.* **28**, 201–204. <https://doi.org/10.1093/nar/28.1.201> (2000).
17. Chapman, S. N. A Molecular Analysis of Potato Virus X. Ph.D. Thesis, Cambridge University (1991).
18. Kavanagh, T. *et al.* Molecular analysis of a resistance-breaking strain of potato virus X. *Virology* **189**, 609–617 (1992).
19. Brown, J. A. Unraveling the structure and biological functions of RNA triple helices. *Wiley Interdiscip. Rev. RNA*. **11**, e1598. <https://doi.org/10.1002/wrna.1598> (2020).
20. Devi, G., Zhou, Y., Zhong, Z., Toh, D. K. & Chen, G. RNA triplexes: From structural principles to biological and biotech applications. *Wiley Interdiscip. Rev. RNA* **6**, 111–128. <https://doi.org/10.1002/wrna.1261> (2015).
21. Chen, I. H., Chu, C. H., Lin, J. W., Hsu, Y. H. & Tsai, C. H. Maintaining the structural integrity of the Bamboo mosaic virus 3' untranslated region is necessary for retaining the catalytic constant for minus-strand RNA synthesis. *Virology* **10**, 208 (2013).
22. Fuentes, S. *et al.* The phylogeography of potato virus X shows the fingerprints of its human vector. *Viruses* **13**, 644–667. <https://doi.org/10.3390/v13040644> (2021).
23. Haasnoot, P. C., Olsthoorn, R. C. L. & Bol, J. F. The Brome mosaic virus subgenomic promoter hairpin is structurally similar to the iron-responsive element and functionally equivalent to the minus-strand core promoter stem-loop C. *RNA* **8**, 110–122. <https://doi.org/10.1017/s1355838202012074> (2002).
24. Koenig, R. *et al.* Nemesia ring necrosis virus: A new tymovirus with a genomic RNA having a histidylatable tobamovirus-like 3' end. *J. Gen. Virol.* **86**, 1827–1833. <https://doi.org/10.1099/vir.0.80916-0> (2005).
25. Miras, M. *et al.* Interfamilial recombination between viruses led to acquisition of a novel translation-enhancing RNA element that allows resistance breaking. *New Phytol.* **202**, 233–246 (2014).
26. Arhab, Y., Mišćicka, A., Pestova, T. V. & Hellen, C. U. T. Horizontal gene transfer as a mechanism for the promiscuous acquisition of distinct classes of IRES by avian caliciviruses. *Nucleic Acids Res.* **50**, 1052–1068. <https://doi.org/10.1093/nar/gkab1243> (2022).
27. Lin, N. S. & Hsu, Y. H. A satellite RNA associated with bamboo mosaic potexvirus. *Virology* **202**, 707–714. <https://doi.org/10.1006/viro.1994.1392> (1994).
28. Huang, Y. W. *et al.* Hsp90 interacts specifically with viral RNA and differentially regulates replication initiation of Bamboo mosaic virus and associated satellite RNA. *PLoS Pathog.* **8**, e1002726. <https://doi.org/10.1371/journal.ppat.1002726> (2012).
29. Girard, G., Gulyaev, A. P. & Olsthoorn, R. C. L. Upstream start codon in segment 4 of North American H2 avian influenza A viruses. *Infect. Genet. Evol.* **11**, 489–495. <https://doi.org/10.1016/j.meegid.2010.12.014> (2011).

## Acknowledgements

We thank Dr. M. Aranda and Dr. R. Sempere for kindly providing the PepMV infectious cDNA clone pTOPO-T7 PepXL6 DNA. We are grateful to Dr. A.P. Gulyaev for advice on pseudoknot structures.

## Author contributions

R.C.L.O. conceived the experiments and wrote the main manuscript text. I.C.L. and C.A.O. participated in the design of the experiments. R.C.L.O. and C.A.O. conducted the experiments, R.C.L.O., I.C.L. and C.A.O. analysed the results. R.C.L.O. and I.C.L. raised funds and provided resources. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15598-5>.

**Correspondence** and requests for materials should be addressed to R.C.L.O.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2022