Similarities between the secondary structure of satellite tobacco mosaic virus and tobamovirus RNAs

Alexander P. Gultyaev,1,2† Eke van Batenburg1 and Cornelis W. A. Pleij2*

1 Institute of Theoretical Biology, University of Leiden, Kaiserstraat 63, 2311 GP Leiden and 2 Department of Biochemistry, Gorlaeus Laboratories, University of Leiden, P.O. Box 9502, 2300 RA Leiden, The Netherlands

The secondary structure of satellite tobacco mosaic virus (STMV) RNA was predicted using computer simulations of RNA folding. The analogies of structural elements in the 3' end untranslated regions (3'-UTR) of tobamoviral RNAs were analysed. In addition to the tRNA-like structure and pseudoknot stalk, which are found in all known RNAs of tobamoviruses and STMV, another region of stable consecutive pseudoknots was predicted in the 3'-UTR of STMV RNA. A similar pattern of repeated structural units, containing pseudoknot stalks and parts of the tRNA-like structure, was also found in odontoglossum ringspot virus (ORSV) RNA 3'-UTR. The predictions on the structure are supported by sequence comparisons which point to an important functional role of 3' terminal pseudoknots in STMV RNA as well as in other tobamoviral RNAs. The possible participation of pseudoknotted structures in the interactions with coat protein in STMV is discussed.

Satellite tobacco mosaic virus (STMV) is the natural satellite of the helper tobacco mild green mosaic virus (TMGMV, U2/U5-TMV strain). Its genome consists of ssRNA of 1059 nucleotides (Mirkov et al., 1989). STMV is also maintained by other tobamoviruses, including tobacco mosaic virus (U1-TMV), although these helpers are less efficient. The 240 3'-terminal nucleotides of STMV RNA exhibit considerable sequence similarity with the Y-terminal part of TMGMV and U1-TMV RNAs. However, such a similarity was not seen (Mirkov et al., 1989) between the coding regions of STMV and tobamoviral RNAs. Also, in contrast to the rod-shaped structure of TMV, STMV has a spherical morphology. The icosahedral structure of STMV has been deduced using X-ray data (Larson et al., 1993a, b). The X-ray data also allowed some characteristics of the RNA that bind to protein in the virion to be determined. Each protein dimer was shown to interact with a dsRNA segment of seven base-pairs with a stacked base at each 3'-end. A secondary structure for STMV RNA has been proposed (Kurath et al., 1993; Larson et al., 1993b) on the basis of computer-aided folding data, suggesting the presence of RNA structural elements as seen by X-ray diffraction data. However, it is possible to suggest alternative foldings that allow comparisons to be made with RNAs of helper viruses.

Here we present the results of RNA structure predictions combined with a comparative analysis of structural elements and these reveal considerable structural similarities between tobamoviral RNAs, including regions with low sequence identity.

The RNA structure calculations were performed using algorithms of folding simulation, taking into account pseudoknot formation (Abrahams et al., 1990; Gultyaev, 1991). A modified version of the STAR program (Abrahams et al., 1990) was also used, which incorporated an algorithm that employed Monte-Carlo simulation of RNA folding (Gultyaev, 1991).

Part of the predicted structure of STMV RNA, comprising 406 nucleotides (654 to 1059) in the 3'-end-untranslated region (3'-UTR) is shown in Fig. 1(a). The most interesting feature of this folding is the close similarity of the structure formed by the last 186 3'-terminal nucleotides to that proposed for the 3'-UTR of tobacco mosaic virus (TMV) and other tobamoviruses (van Belkum et al., 1985; Pleij et al., 1987; Mans et al., 1991), shown in Fig. 1(b). The nucleotides 874 to 1059 at the 3'-end of STMV RNA can be folded in a way which is identical to the TMV analogue. Two characteristic domains may be distinguished in both structures. The first is the tRNA-like structure at the very 3' end of RNA (nucleotides 952 to 1059) and the second is the stretch of three consecutive pseudoknots upstream of it (positions 874 to 950).

The tRNA-like structure, which is present in a number of RNAs from various plant viruses, including tobamoviruses, was shown to be necessary for aminoacylation (for a review see Mans et al., 1991). This structure is very
Fig. 1. Predicted structure of STMV (a) and TMV (b) RNAs.
similar in all known RNA sequences of histidine-accepting tobamoviruses (van Belkum et al., 1985; Pleij et al., 1987; Garcia-Arenal, 1988; Mans et al., 1991) with one exception, RNA from the valine-accepting cowpea strain of TMV. The latter adopts a tRNA-like structure that is closely related to that of TMV RNA (van Belkum et al., 1985; Pleij et al., 1987; Garcia-Arenal, 1988; Isomura et al., 1991). Such a stalk, containing three pseudoknots, is also predicted in STMV RNA (Fig. 1a). This part of the tobamoviral RNA 3'-UTR seems to play an important role in virus replication. It was shown (Takamatsu et al., 1990) that the pseudoknot of TMV RNA (positions 6260 to 6289) with the stem containing an adenine that formed a bulge is very important for viral multiplication. This pseudoknot, with the adjacent one upstream, were shown to regulate viral mRNA translation and a consensus sequence and structural requirements were proposed (Leathers et al., 1993). The fact that STMV RNA can adopt the same folding in this region lends further support to our predictions and suggests an important functional role for the structure of STMV. A similar stalk of pseudoknots was also predicted for brome mosaic virus RNA which belongs to a different virus group (Pleij et al., 1987) and this structure was also shown to be important for genome replication (Lahser et al., 1993).

The closer sequence similarity of the STMV tRNA-like structure to that of TMV rather than to the helper TMGMV has been noted (Solis & Garcia-Arenal, 1990). The pseudoknot stalk region that is located upstream of the tRNA-like structure in STMV RNA (nucleotides 874 to 950) also shows more sequence and structural similarity with TMV RNA than with that of TMGMV. The 3'-proximal pseudoknot (nucleotides 6260 to 6289), which was shown to be functionally important in TMV RNA (Takamatsu et al., 1990), is completely conserved in STMV RNA, even containing the same nucleotide sequence (nucleotides 921 to 950) (Fig. 1).

Five stem–loop structures are predicted in the STMV RNA 3'-UTR region (positions 736 to 871) located upstream of the pseudoknot stalk (Fig. 1a). A similar pattern of consecutive stem–loop elements can also be proposed for the regions upstream of the pseudoknot stalks in other tobamovirus RNAs (not shown). However, in this region it is difficult to derive some consensus structure which would be formed by the homologous nucleotides in all tobamovirus RNAs. Nevertheless, one structural feature is strikingly common in all RNAs studied, namely the stem–loop structure formed by the sequence (for TMV RNA, positions 6182 to 6206, Fig. 1b) surrounding the coat protein gene stop-codon. It should be emphasized that these pairings are formed in spite of differences in the sequences. This may point to the non-randomness of the proposed foldings. In STMV RNA the stem–loop structure predicted upstream (bases 851 to 871) of the pseudoknot stalk (Fig. 1a) seems to resemble that of TMV RNA (Fig. 1b) much more in comparison with analogous RNA foldings of other tobamoviruses, including that of TMGMV. Thus the similarity between TMV and STMV RNA structures, which is seen in the tRNA-like structure and the pseudoknot stalk, may be extended further upstream.

A second stable pseudoknot stalk of three pseudoknots is predicted in the 5'-proximal part of STMV RNA 3'-UTR (bases 654 to 728, Fig. 1a). We could not find any similar folding pattern in TMGMV or TMV RNAs. The 3'-UTR in STMV RNA is much longer (418 nucleotides) than those in the RNAs of its helpers, TMGMV and TMV (210 and 204 nucleotides, respectively). Other tobamoviral RNAs, with the exception of odontoglossum ringspot virus (ORSV), also have 3'-UTRs of approximately 200 nucleotides. The length of ORSV RNA 3'-UTR (414 nucleotides) (Isomura et al., 1991) is very close to that of STMV. The ORSV RNA 3'-UTR, however, contains three repeated sequences, each of them containing two pseudoknots conserved in all tobamoviruses (Isomura et al., 1991).

We have analysed possible foldings of ORSV RNA 3'-UTR. The regions between the sequence repeats containing two pseudoknots mentioned above (nucleotides 1490 to 1546, 1594 to 1650 and 1706 to 1762) can also be folded in a repeating manner (Fig. 2), which is similar to a part of the tRNA-like structure. This results in three structural repeats (nucleotides 1467 to 1593, 1595 to 1704 and 1707 to 1828), each of them containing two pseudoknots from the stalk and the part of tRNA-like structure corresponding to the anti-codon domain downstream of it. The aminoacyl acceptor domain (nucleotides 1829 to 1865) is located downstream of these repeating units, thus forming the entire tRNA-like structure at the very 3'-end of ORSV RNA. In the two 5'-proximal repeats the hairpins analogous to the anti-codon domain are shorter in comparison with that of the 3'-proximal one. They differ from each other by deletion of one A–U base pair and a shortened hairpin loop in the case of the 5'-proximal repeat. The deletion of this base pair strongly supports the proposed folding, as well as covariations (compensatory base changes) in the stems forming the pseudoknotted part of the tRNA-like structure. In each of these pseudoknots, one of the two loops consists of four nucleotides, similar to the structures in other tobamoviruses, whereas the other loop contains mostly A and U residues, which also seems to be a feature that has been conserved.

The pseudoknot stalk in the 5'-proximal repeat
contains three pseudoknots (nucleotides 1467 to 1542) adjacent to the coat protein gene stop-codon (Fig. 2), similar to the folding of STMV RNA 3'-UTR (Fig. 1). In this way, ORSV RNA 3'-UTR forms a structural domain containing 11 consecutive pseudoknots, and it should be emphasized that this striking feature is consistent with sequence and structural comparisons which are usually considered as the best support for RNA structure predictions (Chastain & Tinoco, 1991).

The function of such additional pseudoknot stalks in STMV and ORSV RNAs is unknown. However, it should be mentioned that the locations of these additional stalks in relatively long 3'-UTRs just downstream of coat protein stop-codons are similar to that of

Fig. 2. Predicted structure of ORSV RNA.
the stalks that are found in all other tobamoviruses. Such a position of pseudoknot(s) just downstream of the stop-codon may be a common feature in non-polyadenylated plant viral RNAs (Pleij et al., 1987; ten Dam et al., 1992). The stable pseudoknot stalks were also predicted in the long 3'UTRs of satellite tobacco necrosis viruses (types 1 and 2), and this suggestion was supported by sequence comparisons (Danthinne et al., 1991). It is possible that such long quasi-continuous RNA helices may be important for virus replication (Danthinne et al., 1991; Lahser et al., 1993) and/or mRNA translation (Leathers et al., 1993).

The presence of duplicated sequences and/or structural elements in STMV and ORSV RNA 3'UTRs allows one to speculate about a possible role of recombination events in their evolution. RNA recombination between genome segments of plant viruses and/or their satellite RNAs has been shown (Buijarski & Kaesberg, 1986; Cascone et al., 1990). The mosaic composition of structural elements in the RNA of the cowpea strain of TMV, containing both tobamoviral- and tymoviral-like regions, also indicates a possible acquisition of these structures by recombination (van Belkum et al., 1985).

X-ray data on STMV crystals have revealed multiple protein–RNA contacts in the interior of the virion (Larson et al., 1993a, b). The coat protein binding sites in the RNA were shown to be represented by double-helical segments of seven base-pairs with unpaired stacked nucleotides at their 3'-ends. Some structural features of the protein-bound RNA, namely overwinding of helices and stacked unpaired bases at their 3'-ends (Larson et al., 1993a, b) resemble those found in pseudoknots (Puglisi et al., 1990). These analogies suggest a possible role of pseudoknots in binding to the coat protein dimer.

The coding region of STMV RNA may be folded in several ways (not shown) forming highly-ordered structures. However, the particular features of these foldings remain unknown in the absence of experimental and/or phylogenetic data. Some parts of our predicted secondary structure for STMV RNA appeared to be consistent with those predicted by the search of minimal free energy (Kurath et al., 1993; Larson et al., 1993b).

We believe that the structure proposed here for the 3'-UTR is more reliable than minimum energy predictions (Kurath et al., 1993; Larson et al., 1993b) because it contains structural elements, pseudoknots in particular, that are similar to those predicted in the helper virus RNAs.

Thus it may be concluded that the secondary and tertiary structure of the long 3'-UTR of STMV RNA may be very important for satellite virus reproduction. The analogies between this structure and that of the helper virus genome suggest similarities in their functioning.

This work was supported by grants given to A.P.G. from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek and the European Molecular Biology Organization.

References


DANTHINNE, X., SIEURINK, J., VAN MONTAGU, M., PLEIJ, C. W. A. & BOSCH, L. (1987). Contributions of genome segments of plant viruses and/or their satellite RNAs has been shown (Buijarski & Kaesberg, 1986; Cascone et al., 1990). The mosaic composition of structural elements in the RNA of the cowpea strain of TMV, containing both tobamoviral- and tymoviral-like regions, also indicates a possible acquisition of these structures by recombination (van Belkum et al., 1985).

X-ray data on STMV crystals have revealed multiple protein–RNA contacts in the interior of the virion (Larson et al., 1993a, b). The coat protein binding sites in the RNA were shown to be represented by double-helical segments of seven base-pairs with unpaired stacked nucleotides at their 3'-ends. Some structural features of the protein-bound RNA, namely overwinding of helices and stacked unpaired bases at their 3'-ends (Larson et al., 1993a, b) resemble those found in pseudoknots (Puglisi et al., 1990). These analogies suggest a possible role of pseudoknots in binding to the coat protein dimer.

The coding region of STMV RNA may be folded in several ways (not shown) forming highly-ordered structures. However, the particular features of these foldings remain unknown in the absence of experimental and/or phylogenetic data. Some parts of our predicted secondary structure for STMV RNA appeared to be consistent with those predicted by the search of minimal free energy (Kurath et al., 1993; Larson et al., 1993b).

We believe that the structure proposed here for the 3'-UTR is more reliable than minimum energy predictions (Kurath et al., 1993; Larson et al., 1993b) because it contains structural elements, pseudoknots in particular, that are similar to those predicted in the helper virus RNAs.

Thus it may be concluded that the secondary and tertiary structure of the long 3'-UTR of STMV RNA may be very important for satellite virus reproduction. The analogies between this structure and that of the helper virus genome suggest similarities in their functioning.

This work was supported by grants given to A.P.G. from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek and the European Molecular Biology Organization.


(Received 25 February 1994; Accepted 17 May 1994)