Sequence homology between potato spindle tuber viroid and U3B snRNA

Tamás Kiss, János Pósfai+ and Ferenc Solymosy*

Institute of Plant Physiology and +*Institute of Biophysics, Biological Research Center, Hungarian Academy of Sciences, PO Box 521, Szeged H-6701, Hungary*

Received 26 September 1983

Sequence homology was found by computer analysis between potato spindle tuber viroid (PSTV) RNA and U3B snRNA of Novikoff hepatoma cells. This homology is colinear in arrangement, extends in length to 81% of the entire U3B snRNA molecule and is involved in the PSTV molecule unique sites which, if depicted in terms of the secondary structure of the circular PSTV molecule, reveal a conspicuous regularity in their location. A strong relation in primary structure between PSTV and U3B snRNA is demonstrated by statistical analysis.

PSTV *U3B snRNA* *Sequence homology*

1. INTRODUCTION

Viroids [1], a distinct class of plant pathogenic agents, are covalently closed circular single-stranded RNA molecules [2] with a particular, DNA-like secondary structure [3]. The complete nucleotide sequence and secondary structure of some of them are known [2,4–8].

Eukaryotic small nuclear RNAs (snRNAs), of a largely unknown function, are linear single-stranded RNA molecules with a hairpin-like secondary structure (cf. [9]). All of the snRNAs of the U series (U1–U6) have been sequenced and shown to contain a cap structure at their 5′-end and modified bases (cf. [9]). They are very conserved molecular species in nature, from dinoflagellates to man (cf. [9]). In plants no U-type snRNAs have been identified so far.

The following features, common to U snRNAs and viroids, point to a formal (possibly phylogenetic) relationship between these two molecular entities:

(i) Both have been found so far only in eukaryotes;

(ii) Both are localized and synthesized (replicated) in the cell nucleus (cf. [9] for U snRNAs and [10–12] for viroids);

(iii) In the synthesis (replication) of both of them an enzyme whose sensitivity to α-amanitin corresponds to that of DNA-dependent RNA polymerase II is involved ([13,14] for U1, U2 and U3 snRNAs and [15–17] for viroids);

(iv) Both are small single-stranded RNA molecules comprising 107–214 (cf. [9]) and 246–371 (cf. [18]) nucleotide residues, depending on the individual species of U snRNA and viroid, respectively;

(v) Both seem to be devoid of tRNA and mRNA activities in the appropriate in vitro systems ([19] for snRNAs and [20–22] for viroids);

(vi) Avocado sunblotch viroid (ASBV) contains a number of sequences homologous with stretches of U5 snRNA from Novikoff hepatoma cells [23];

(vii) All viroids tested so far contain one [6,8,23,24] or more [23,25] stretches in their molecule that show partial homology with a stretch at the 5′-end of U1 snRNA of vertebrate origin, supposed to play a vital role in the splicing of pre-mRNA [26,27]. This

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.

observation has given rise to speculations that viroids may exert their pathogenic action by interfering with the normal way of processing mRNA [6,24,25].

Circularity of viroids *vs* linearity of U snRNAs as well as the absence of modified bases and a cap structure in viroids *vs* their presence in U snRNAs are of limited value as arguments against a possible phylogenetic relationship of the two, because these differences might have arisen later in evolution.

We here show that there is a striking sequence homology between U3B snRNA from Novikoff hepatoma cells and unique stretches of the PSTV molecule.

2. MATERIALS AND METHODS

The nucleotide sequences of PSTV, U snRNAs of Novikoff hepatoma cells, D2 snRNA and Intron I in the 16 S–23 S spacer region from tobacco chloroplasts were taken from [6], [9], [28] and [29], respectively.

To detect possible sequence homologies between PSTV and U snRNAs a computer analysis based on a program described in [30] was used.

To assess the significance of the sequence homology found between PSTV and U3B snRNA, the Tinoco matrices [31] of the aligned sequences of PSTV, U3B snRNA and a number of controls were formed as in [32]. The matrix, representing with its points matches between the sequences, was divided into equal boxes. The number of boxes containing N or less scores was plotted against N. Homogeneity tests, using χ^2 statistics, were performed on pairs of such cumulative frequency distributions derived from the investigated sequences and from random patterns with base compositions of the non-random ones.

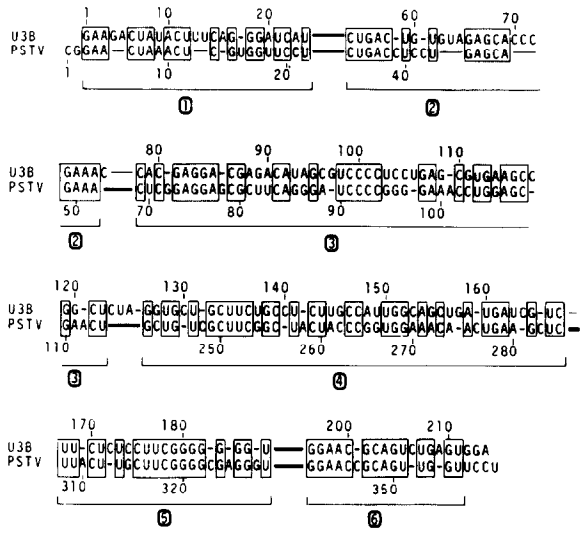

Fig.1. Sequence homologies (boxed areas) between regions of the 214 nucleotide-residue-long Novikoff hepatoma U3B snRNA molecule (U3B) and those (1–6 as indicated) of the 359 nucleotide-residue-long PSTV molecule (PSTV). Sequences are aligned for maximum homology. The figures above and below the characters refer to the positions of nucleotide residues in the linear U3B snRNA (modified nucleotides not marked and ψ indicated as U) and to those in the circular PSTV (conventional numbering system as introduced by [2]), respectively. Non-homologous regions are indicated by bold-face lines in the respective sequence. Thin lines mean continuity in sequences.

3. RESULTS AND DISCUSSION

Screening of the extent of sequence homologies between PSTV on the one hand and the 6 U snRNAs (U1–U6) on the other, showed that it is the U3B species of the U3 snRNA molecule (the only U snRNA of nucleolar origin) which contains the maximum number of the longest stretches that

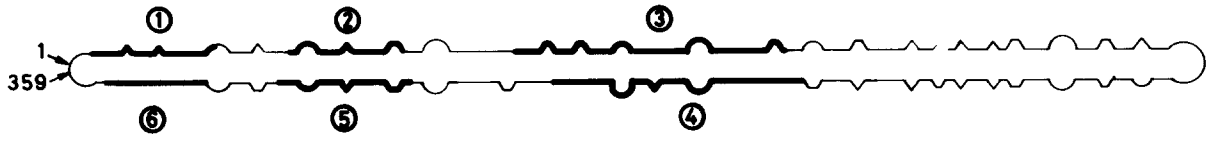

Fig.2. Schematic diagram of the secondary structure of PSTV (after [2]). The regions (1–6 as in fig.1) that show sequence homologies with stretches of U3B snRNA are drawn in bold-face. The figures with the arrows indicate the first and last nucleotide residues in the circular molecule according to the conventional numbering system [2].

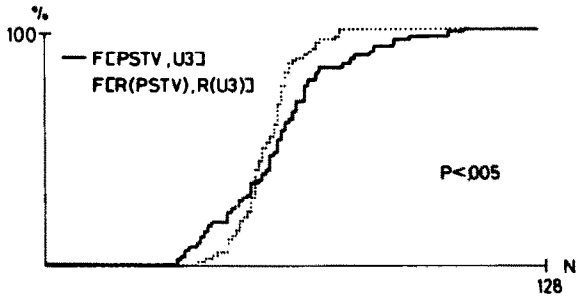

Fig.3. Frequency functions showing the spatial distributions of points of Tinoco matrices derived from PSTV *vs* U3B snRNA (——) and from random patterns (– – –). Size of matrix 359 × 216, order of matches 2, size of boxes 30 × 30. Abscissa, *N*; ordinate, % of boxes containing *N* or less points; R (NAME), random pattern with base composition of sequence NAME; F [NAME1, NAME2], distribution function from the Tinoco matrix of sequences NAME1, NAME2; *P*, the probability that the two frequency functions have the same distribution.

are located in a colinear way with respect to homologous stretches in the PSTV molecule.

Fig.1 shows the sequence homology (boxed areas) between PSTV and U3B snRNA. This homology:

(i) Is colinear in arrangement;

(ii) Extends in length to 81% (174 nucleotide residues) of the entire U3B snRNA molecule and 68% (119 nucleotide residues) of this region is actually involved in the formation of bona fide homologies with PSTV;

(iii) Involves in the PSTV molecule unique sites which, if depicted in terms of the secondary structure of the circular PSTV, reveal a conspicuous regularity in their location (fig.2).

A strong relation in primary structure between PSTV U3B snRNA is demonstrated in fig.3 and 4.

Fig.3 shows that the spatial distribution of the points in the Tinoco matrix of PSTV and U3B snRNA differs significantly from that of random patterns. The same low threshold of probability of random occurrence ($p < 0.005$) is obtained if 100% homology is guaranteed (fig.4A) or if two RNA molecules of established sequence homology are compared (fig.4D). On the contrary, a comparison of Tinoco matrices from random, evenly distribu-

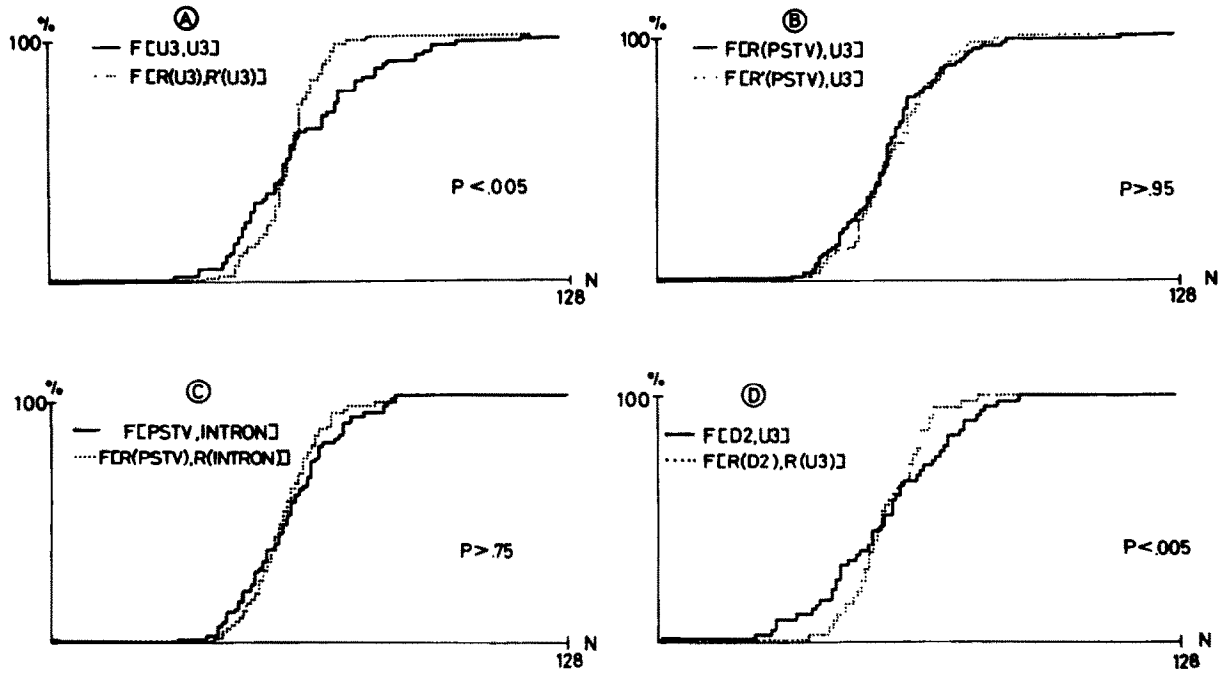

Fig.4. Comparison between the frequency functions showing the spatial distributions of points of Tinoco matrices derived from pairs of sequences and their randomized derivatives. (A) U3B snRNA *vs* U3B snRNA; (B) randomized PSTV *vs* U3B snRNA; (C) PSTV *vs* a 216 nucleotide-residue-long portion of Intron I [29]; (D) D2 snRNA *vs* U3 snRNA. Notations and parameters as in fig.3.

ted patterns (fig.4B) or from unrelated natural sequences (fig.4C) gives incomparably higher p-values ($p > 0.95$ and $p > 0.75$, respectively).

The sequence homology found between PSTV and U3B snRNA may be the reflection of a phylogenetic process by which viroids evolved. Such an evolution might have taken place either at the transcriptional level (transcription from a U3 snRNA pseudogene) or at the post-transcriptional level (e.g., circularization of U3 snRNA and fusion with some other small circularized RNA molecule).

We suggest that the left half of the circular PSTV is phylogenetically related to U3 snRNA. Whether this U3 snRNA was an alleged plant equivalent of mammalian U3 snRNA or whether it originated from some eukaryote other than plants, cannot be decided at present. Isolation and sequencing of the putative plant snRNAs, their genes and pseudogenes, may yield more evidence to answer this question.

The remote possibility, however, cannot be ruled out that sequence homology between a viroid and an snRNA may just be a reflection of some common structural requirement for biosynthesis and thus be the result of structural convergence.

As far as functional implications of our finding are concerned, it is tempting to speculate whether viroids, owing to structural similarities with snRNAs, such as those reported here, may interfere with any of those nuclear events in which snRNAs are involved, whatever these events are.

ACKNOWLEDGEMENT

This work has been supported by grant no. 363–82–1.6 of the Hungarian Academy of Sciences.

REFERENCES

[1] Diener, T.O. (1971) Virology 45, 411–428.
[2] Gross, H.J., Domdey, H., Lossow, C., Jank, P., Raba, M., Alberty, H. and Sänger, H.L. (1978) Nature 273, 203–208.
[3] Riesner, D., Henco, K., Rokohl, U., Klotz, G., Kleinschmidt, A.K., Gross, H.J., Domdey, H. and Sänger, H.L. (1979) J. Mol. Biol. 133, 85–115.
[4] Haseloff, J. and Symons, R.H. (1981) Nucleic Acids Res. 9, 2741–2752.
[5] Visvader, J.E., Gould, A.R., Bruening, G.E. and Symons, R.H. (1982) FEBS Lett. 137, 288–292.
[6] Gross, H.J., Krupp, G., Domdey, H., Raba, M., Jank, P., Lossow, C., Alberty, H., Ramm, K. and Sänger, H.L. (1982) Eur. J. Biochem. 121, 249–257.
[7] Symons, R.H. (1981) Nucleic Acids Res. 9, 6527–6537.
[8] Haseloff, J., Mohamed, N.A. and Symons, R.H. (1982) Nature 299, 316–322.
[9] Busch, H., Reddy, R., Rothblum, L. and Choi, Y.C. (1982) Annu. Rev. Biochem. 51, 617–654.
[10] Diener, T.O. (1971) Virology 43, 75–89.
[11] Takahashi, T. and Diener, T.O. (1975) Virology 64, 106–114.
[12] Semancik, J.S., Tsuruda, D., Zaner, L., Geelen, J.L.M.C. and Weathers, J.G. (1976) Virology 69, 669–676.
[13] Frederiksen, S., Hellung-Larsen, P. and Gram-Jensen, E. (1978) FEBS Lett. 87, 227–231.
[14] Gram-Jensen, E., Hellung-Larsen, P. and Frederiksen, S. (1979) Nucleic Acids Res. 6, 321–330.
[15] Mühlbach, H.-P. and Sänger, H.L. (1979) Nature 278, 185–188.
[16] Rackwitz, H.R., Rohde, W. and Sänger, H.L. (1981) Nature 291, 297–301.
[17] Flores, R. and Semancik, J.S. (1982) Proc. Natl. Acad. Sci. USA 79, 6285–6288.
[18] Sänger, H.L. (1982) in: Encyclopedia of Plant Physiology (Parthier, B. and Boulter, D. eds) vol.14B, pp.368–454, Springer, Berlin.
[19] Hellung-Larsen, P. (1977) Low Molecular Weight RNA Components in Eukaryotic Cells, p.117, FADL's Forlag, Copenhagen.
[20] Davies, J.W., Kaesberg, P. and Diener, T.O. (1974) Virology 61, 281–286.
[21] Hall, T.C., Wepprich, R.K., Davies, J.W., Weathers, L.G. and Semancik, J.S. (1974) Virology 61, 486–492.
[22] Semancik, J.S., Conejero, V. and Gerhart, J. (1977) Virology 80, 218–221.
[23] Kiss, T. and Solymosy, F. (1982) FEBS Lett. 144, 318–320.
[24] Diener, T.O. (1981) Proc. Natl. Acad. Sci. USA 78, 5014–5015.
[25] Dickson, E. (1981) Virology 115, 216–221.
[26] Roberts, R. (1980) Nature 283, 132–133.
[27] Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Nature 283, 220–224.
[28] Wise, J.A. and Weiner, A.M. (1980) Cell 22, 109–118.
[29] Takaiwa, F. and Sugiura, M. (1982) Nucleic Acids Res. 10, 2665–2676.
[30] Goad, W.B. and Kanehisa, M.I. (1982) Nucleic Acids Res. 10, 247–263.
[31] Tinoco, I. jr, Uhlenbeck, O.C. and Levine, M.D. (1971) Nature 230, 362–367.
[32] Novotny, J. (1982) Nucleic Acids Res. 10, 127–131.