FEBS Letters 350 (1994) 5–8

FEBS 14335

Complete nucleotide sequence and genome organization of a tobamovirus infecting cruciferae plants

Yu.L. Dorokhov[a], P.A. Ivanov[c], V.K. Novikov[b], A.A. Agranovsky[b], S.Yu. Morozov[a], V.A. Efimov[c], R. Casper[d], J.G. Atabekov[b,*]

[a]*A.N.Belozersky Institute of Physico-Chemical Biology and* [b]*Department of Virology, Moscow State University, Moscow 119899, Russian Federation*
[c]*M.M.Shemiakin Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow 117871, Russian Federation*
[d]*Federal Biological Research Center for Agriculture and Forestry, Institute of Biochemistry and Plant Virology, Messeweg 11/12, D-3300 Braunschweig, Germany*

Received 13 June 1994

Abstract

Genomic RNA sequence of a tobamovirus infecting cruciferae plants (cr-TMV) was determined. The RNA is composed of 6312 nucleotides and contains four ORFs encoding the proteins of 122K (ORF1), 178K (ORF2), 29K (ORF3) and 18K (capsid protein, ORF4). ORF4 overlaps ORF3 by 74 nucleotides and the overlapping region can be folded into a stable hairpin structure. The 3'-terminal region of the cr-TMV RNA preceding the tRNA-like structure was shown to form six potentially stable pseudoknots.

Key words: cDNA cloning; Nucleotide sequence; Amino acid sequence homology; Plant positive-strand RNA virus; Tobacco mosaic virus

1. Introduction

Tobamoviruses represent a group of rod-shaped plant viruses with undivided positive-sense RNA genomes of approximately 6400 nucleotides. Tobamovirus RNA encodes at least four proteins. The products of direct translation of the genomic RNA (130K and its readthrough 180K product) are involved in replication of viral RNA [1]. The 30K protein, which is required for cell-to-cell movement [2,3] and the 17K CP are translated from individual subgenomic mRNAs [4].

Recently a crucifer-infecting tobamovirus (cr-TMV) has been isolated in our laboratory from *Olearacia officinalis* L. The virus showed distant serological relationship to TMV U1 and some other tobamoviruses, and infected systemically the members of the Cruciferaceae family (*Brassica chinensis* L., *B. rapa* L., *B. napus* L., *B. compestris* L. and *A. thaliana*) as well as *Nicotiana tabacum* L. Special interest to cr-TMV is dictated by its ability to infect *A. thaliana*, the species being a perspective model for studying the virus-plant interactions.

In this work we determined the complete nucleotide sequence of the cr-TMV genomic RNA and compared it with those of other tobamoviruses.

*Corresponding author. Fax: (9) (095) 938-0601;
E-mail: libro@genebee.msu.su

Abbreviations: K, kilodalton; ORF, open reading frame; TP, transport protein; CP, coat protein; TMV, tobacco mosaic virus; M_r, relative molecular weight.

2. Materials and methods

2.1. Virus purification and RNA isolation

Turnip plants (*B. rapa* L.) were used for the virus propagation. Two weeks after inoculation, the leaves were harvested and stored at −60°C. The virus was purified according to the method reported earlier [5]. RNA was isolated from the purified virus preparation by phenol-SDS extraction, dissolved in water and kept at −70°C until use.

2.2. cDNA synthesis and cloning

The cDNA synthesis system (Promega) based on the method of Gubler and Hoffman [6], was used to prepare the double-stranded blunt-ended cDNA from total cr-TMV RNA or the 3'polyadenylated RNA primed by random primer or oligo(dT), respectively. The cDNA were inserted into *Sma*I-digested plasmid vectors pBS (Stratagene) or pGEM-3 (Promega) followed by transformation of competent *E. coli* XL-1 cells.

2.3. DNA sequencing and sequence analysis

Sequencing of cDNA inserts was performed by the dideoxynucleotide chain termination method [7] using Sequenase kits (USB). Sequence data were analysed using the GENEBEE package [8].

3. Results and discussion

The first set of cDNA clones used in this work was obtained using random priming of total cr-TMV RNA. Computer sequence analysis revealed the cDNA clones having obvious sequence similarity to the 130/180K, 30K and coat protein genes. A region between nt 4877 to 5809 where differences between the sequences of individual cDNA clones were encountered, was amplified by the PCR, cloned and sequenced. The cDNA clones corresponding to the 3'-terminal part of the genomic RNA were obtained using the polyadenylated cr-TMV genomic RNA and oligo(dT) primer. To obtain the

SSDI 0014-5793(94)00721-7

cDNA clones corresponding to internal part of the genome, the specific primers complementary to the regions 1654–1674, 2593–2613, 3913–3933, 4918–4938, respectively, were used. The cr-TMV sequence was determined completely on both strands of cDNA with the exception of the 5′-terminal 125-nt-long region sequenced directly on theviral RNA by chain termination method with reverse transcriptase.

The complete sequence of 6312 nt of the cr-TMV genomic RNA is shown in Fig. 1. The first AUG codon in the sequence (nt 69–71) has optimal nucleotide context [9] and is likely an initiating codon for the ORF1 encoding a protein composed of 1107 amino acids (122K). The second AUG codon in ORF1 (nt 96–98) has a suboptimal nucleotide context. ORF1 is terminated by the amber UAG codon (nt 3381–3383) followed by the CAA and UUA codons, which are typical for the other tobamovirus sequences and are thought to enhance the readthrough of the leaky terminator in ORF1. The ORF2 encoding the putative read-through protein of 1601 amino acids (178K) terminates at residues 4873–4875. In accordance with the sequence data, two major polypeptides with apparent M_r of 122K and 178K were detected among the translation products of cr-TMV RNA in rabbit reticulocyte lysates (data not shown).

ORF3 (nt 4877–5677) initiates 2 bases downstream from the termination codon of ORF2 and encodes a protein of 267 amino acids (29K). This ORF overlaps with the beginning of the coat protein gene for 74 nt (excluding terminator codon). It is worth mentioning that the ORF3/ORF4 overlapping region of cr-TMV, which unusually long for a tobamovirus, can be folded

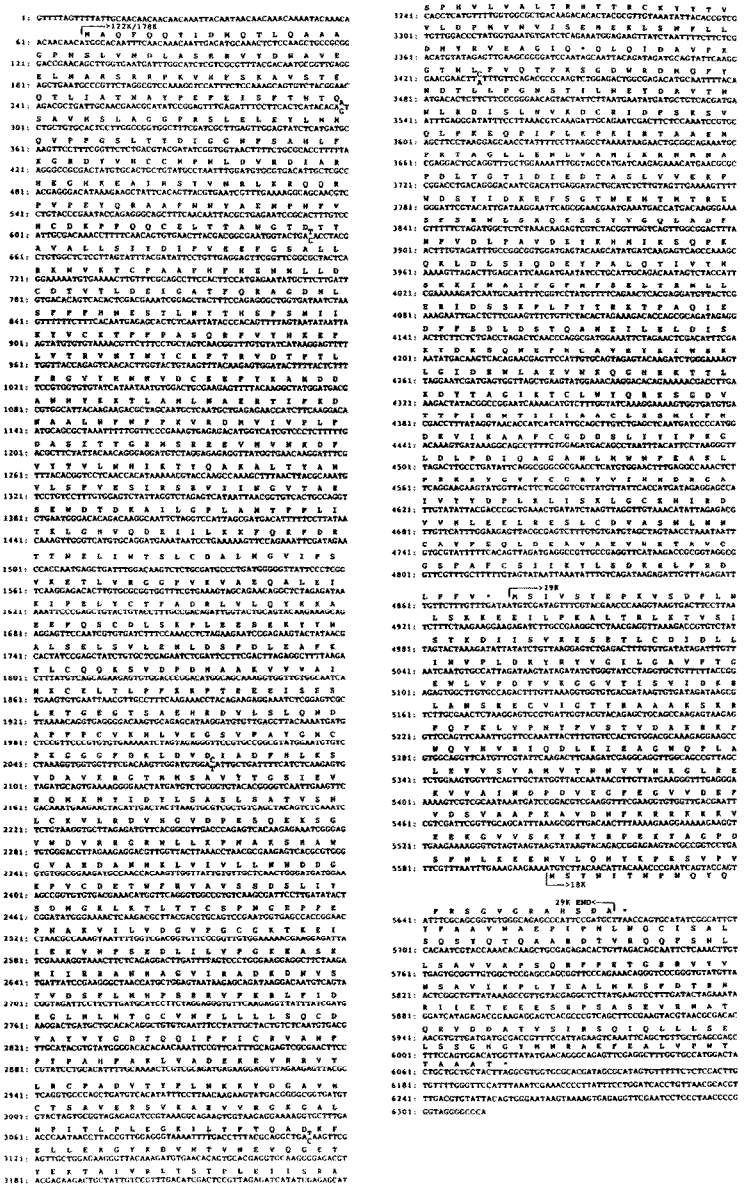

Fig. 1. Complete nucleotide sequence and the deduced amino sequences of the cr-TMV RNA. Molecular weights of the gene products are indicated.

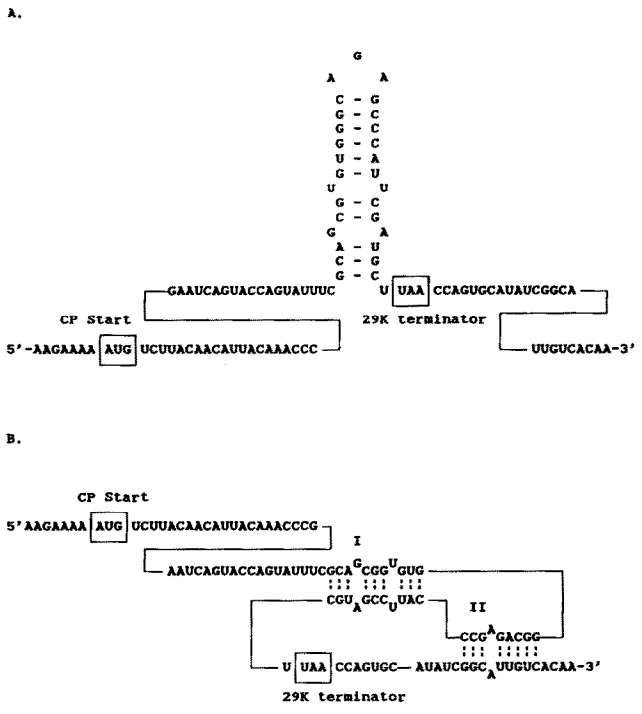

Fig. 2. The tentative foldings of cr-TMV RNA in the overlapping region of cr-TMV 29K and coat protein genes without (A) or with pseudoknot (B).

in a stable hairpin structure (Fig. 2A) located just upstream of the UAA termination codon of the 29K gene. Fig. 2B shows an alternative RNA folding with pseudoknot formation in this region (stems I and II have free energy −5,5 and −4,0 kcal/mol, respectively). Both tentative folding have similar calculated free energy parameters.

The cr-TMV 29K protein shows clear sequence similarity to the transport proteins (TPs) of tobamoviruses (Fig. 3). However, it contains essentially larger proportion of basic amino acids than any other known tobamovirus TP, reflected in its unusually high isoelectric point (p*I* 9.76).

Basing on the five aligned tobamovirus TP sequences, Saito et al. [10] identified two well-conserved regions in their middle portion (I and II) and three regions of marginal conservation in the C-terminal portion (designated A, B and C). Our amino acid sequence comparison (Fig. 3) including ten tobamovirus TP sequences currently available, supports the existence of the conserved domains I and II and shows that the latter region may be considerably more extended than proposed previously [10]. On the other hand, numerous substitutions found in the C-terminal parts of the newly sequenced TP genes, including the 29K ORF of cr-TMV, argue against the conservation of the A, B and C regions (Fig. 3). The region II was found to be involved in the single-stranded nucleic acid-binding ability of tobamoviral TPs identified in TMV-U1 [11] and cr-TMV [12]. The second (C-terminal) portion of the putative RNA binding domains in TMV-U1TP is enriched in positively charged residues [11]. The respective region of the 29K of cr-TMV shows the longest uninterrupted cluster of positively charged amino acids, Lys-Arg-Arg-Lys-Lys-Lys, among the tobamovirus TPs so far sequenced (Fig. 3).

The region of overlap between the TP and CP genes in cr-TMV RNA is considerably longer than that in the other tobamoviruses. Recently it has been shown that the genomes of the subgroup 1a tobamoviruses contain a short ORF (ORF-X) overlapping the 30K and coat

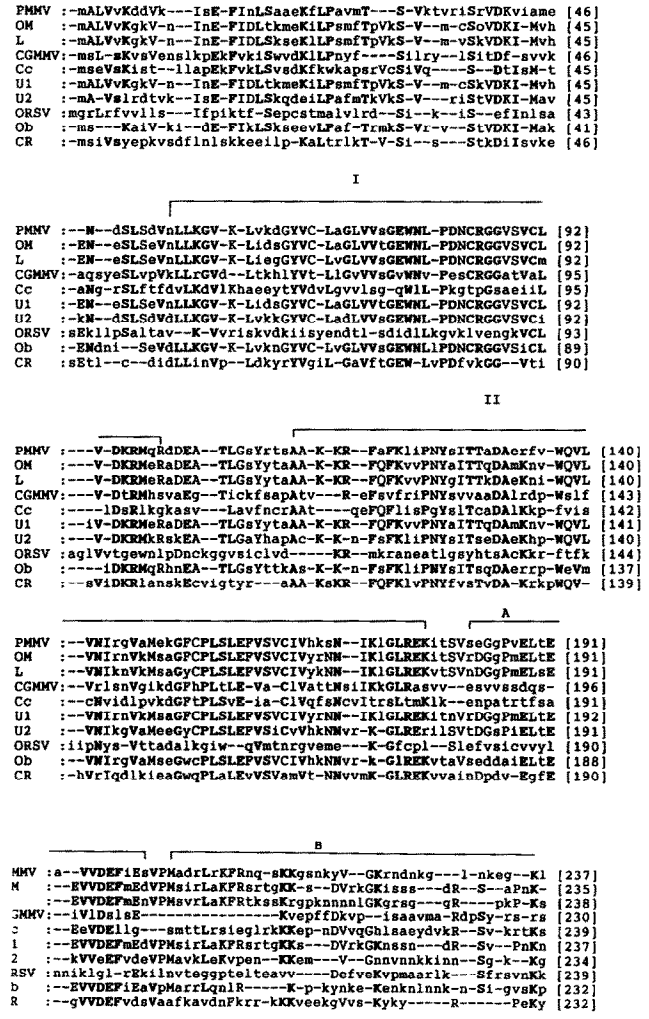

Fig. 3. Alignment of the cr-TMV 29K protein with the 30K TPs of nine tobamoviruses: PMMV, pepper mild mottle virus [14]; TMV strains, OM [10], L [15], Cc (cowpea strain) [16], U1 [17], U2 (tobacco mild green mosaic virus) [18], ORSV, odontoglossum ringspot virus [19] and Ob [20]. Bold capital letters indicate amino acid residues conserved in at least seven sequences. Capital letters show the residues conserved in at least five sequences. Non-conserved residues are shown in lowercase.

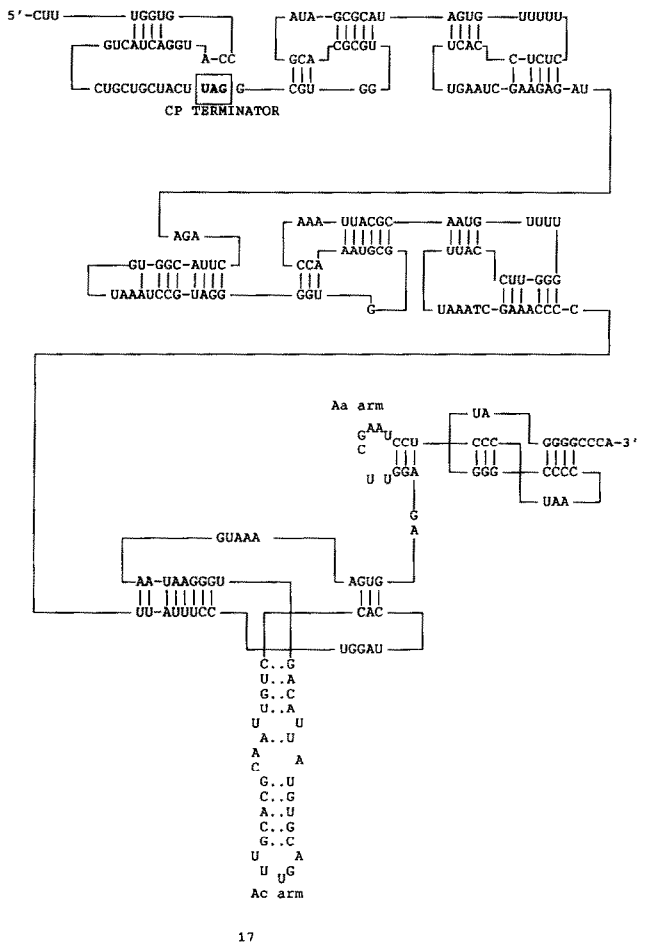

Fig. 4. Proposed pseudoknotted folding of the 3'-terminal portion of cr-TMV RNA including tRNA-like structure. Aa and Ac indicate the aminoacyl acceptor arm and anticodon arm, respectively.

protein genes and encoding a small positively charged protein. When translated from the T7 transcripts, the ORF-X product forms stable specific complex with a yet unidentified protein from rabbit reticulocyte lysates and wheat embryo extracts [13]. The alignment presented below shows a significant similarity between the sequence of 24 C-terminal amino acids of the cr-TMV 29K protein, and the respective region of the TMV-U1 ORF-X protein:

```
TMV-U1 ORF-X       : VLQyHY-siSVrV1visVGRpnrvn
cr-TMV 29K protein: VLQ-HYkpeSVpVfrsgVGRahsda
```

This similarity is apparently higher than that between the C-terminal regions of the tobamoviral TPs (Fig. 3). If the putative ORF-X protein indeed has a specific function upon tobamoviral infection, it should be assumed that its conserved sequence may mediate this function either when expressed as a separate protein (U1-TMV) or as the TP-ORF-X fusion (cr-TMV).

The 3'-terminal non-translated region in the cr-TMV genome can be folded into seven potentially stable pseudoknots and one additonal pseudoknot at the 3'-end of the CP gene compared to only five pseudoknots found in the U1-TMV sequence (Fig. 4). The strongest structural conservation between the cr-TMV and U1-TMV covers the two 3'-proximal pseudoknots forming a tRNA-like structure (Fig. 4). The cr-TMV aminoacylation arm shows the UAAGCUU sequence typical of the tobamovirus RNAs accepting histidine. However, its putative anticodon loop contains the UUG glutamine anticodon instead of the histidine anticodon conserved in several other tobamoviruses. It remains to be determined whether this structural deviation in the cr-TMV RNA influences its esterification with a specific amino acid.

Acknowledgements: This work was supported in part by the Russian Fund for Basic Researches and the International Science Foundation.

References

[1] Ishikawa, M., Meshi, T., Motoyoshi, F., Takamatsu, N. and Okada, Y. (1986) Nucleic Acids Res. 14, 8291–8305.
[2] Meshi, T., Watanabe, Y., Saito, T., Sugimoto, A., Maeda, T. and Okada, Y. (1987). EMBO J. 6, 2557–2563.
[3] Deom, C.M., Show, M.J. and Beachy, R.N. (1987) Science 327, 389–394.
[4] Beachy, R.N. and Zaitlin, M. (1977) Virology 81, 160–169.
[5] Novikov, V.K. and Atabekov, J.G. (1970) Virology 41, 101–107.
[6] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
[7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[8] Brodsky, L.I., Drachev, A.L., Tatuzov, R.L. and Chumakov, K.M. (1990) Biopolimery i Kletka 7,10–14 (in Russian).
[9] Kozak, M. (1991) J. Biol. Chem. 266, 19867–19870.
[10] Saito, T., Imai, Y., Meshi, T. and Okada, Y. (1988) Virology 167, 653–656.
[11] Citovsky, V., Wong, M.L., Show, A., Prasad, B.V.V. and Zambriski, P. (1992) PLant Cell 4,397–411.
[12] Ivanov, K.I., Ivanov, P.A., Timofeeva, K.K., Dorokhov, Y.L. and Atabekov, J.G. (1994) FEBS Lett. (in press).
[13] Morozov, S.Y., Denisenko, O.N., Zelenina, D.A., Fedorkin, O.N., Solovyev, A.G., Maiss, E., Casper, R. and Atabekov, J.G. (1993) Biochimie 75, 659–665.
[14] Alonso, E., Garcia-Luque, L., de la Cruz, A., Wicke, B., Avila-Rincon, M.J., Serra, M.T., Castresana, C. and Diaz-Ruis, J.R. (1991) J. Gen. Virol. 72, 2875–2884.
[15] Ohno, T., Aoyagi, M., Yamanashi, Y., Saito, H., Ikawa, S., Meshi, T. and Okada, Y. (1984) J. Biochem. 96, 1915–1923.
[16] Meshi T., Ohno T. and Okada Y. (1982) Nucleic Acids Res. 10, 6111–6117.
[17] Goelet, P., Lomonossoff, G., Butler, P.J.G., Akam, M., Gait, M.J. and Karn, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5818–5822.
[18] Ugaki, M., Tomiyama, Kakutani, T., Hidaka, S., Kiguchi, T., Nagata, R., Sato, T., Motoyoshi, F. and Nishiguchi, M. (1991) J. Gen. Virol. 72, 1487–1495.
[19] Isomura, Y., Matumoto, Y., Murayama, A., Chatani, M., Inouye, N. and Ikegami, M. (1990) Nucleic Acids Res. 18, 7448.
[20] Ikeda, R., Watanabe, E., Watanabe, Y. and Okada, Y. (1993) J. Gen. Virol. 74, 1939–1944.