Sequence of an oligonucleotide derived from the 3' end of each of the four brome mosaic viral RNAs

(AMINOACYLATION/tRNA-LIKE STRUCTURE)

RANJIT DASGUPTA AND PAUL KAESBERG

Biophysics Laboratory of the Graduate School and Biochemistry Department of the College of Agricultural and Life Sciences, University of Wisconsin, Madison, Wisconsin 53706

Communicated by Heinz Fraenkel-Conrat, August 29, 1977

ABSTRACT A 3'-terminal oligonucleotide fragment, 161 bases long, can be obtained from each of the four brome mosaic virus RNAs by means of nuclease digestion. Like the four intact brome mosaic virus RNAs, each fragment accepts tyrosine in a reaction catalyzed by wheat germ aminoacyl-tRNA synthetase. The complete nucleotide sequence of the RNA 4 fragment has been determined by use of standard radiochemical methods. Comparative data for the fragments from RNAs 1, 2, and 3 show that they have nearly the same sequence as the RNA 4 fragment. The eight bases adjacent to the 3' terminus of the RNA 4 fragment are identical in sequence to the eight terminal bases of tyrosine tRNA from *Torula utilis* and eleven interior bases are identical in sequence to eleven bases encompassing the anticodon region of tyrosine tRNA from *Saccharomyces cerevisiae*, *T. utilis*, and *Escherichia coli*. Nevertheless, reasonable base-pairing schemes yield, at best, a distorted cloverleaf secondary structure.

Nucleotide sequence analysis of the extremities of plant viral nucleic acids has provided substantial information regarding their structure and their function as messengers. For example, the 5' termini of some of these RNAs, in common with most eukaryotic messengers, contain the "cap" structure m'GpppGp. (For an extensive review, see ref. 1.) The efficiency of translation is dependent on the presence of cap (2-4). For brome mosaic virus (BMV) RNA 4, the monocistrionic messenger for BMV coat protein, the initiation codon is only 10 nucleotides from the capped 5' end and this short sequence, together with the cap, constitutes an effective ribosome binding site (5).

The 3' termini of some plant viral RNAs have the unique property among messengers that they can quantitatively accept an amino acid in a reaction catalyzed by aminoacyl-tRNA synthetase. Thus, turnip yellow mosaic virus RNA (6) and egg plant mosaic virus RNA (7) can be charged with valine; tobacco mosaic virus RNA can be charged with histidine (8). Each of the four BMV RNAs can be charged with tyrosine (9). Some picornaviral RNAs can be charged, albeit inefficiently and probably only after they have been fragmented (10, 11). The property of chargeability implies that these RNAs have a tRNA-like structure and possibly also a tRNA-like function (12). A secondary structure, somewhat like that of tRNA, is compatible with the sequences determined for the 3' ends of RNAs with the terminal region of tobacco mosaic virus RNA seems less amenable to folding into a cloverleaf secondary structure (15).

Although the BMV RNAs can be aminoacylated with tyrosine, this amino acid is not donated to nascent peptides upon in vitro translation (16). However, integrity of the 3' end is necessary for infectivity of BMV RNA (17). Thus, like their 5' ends, the 3' ends of these viral RNAs have a distinctive structure and presumably an important, although unknown, function.

We have reported that partial hydrolysis of each of the four BMV RNAs with RNase T1 releases a 3'-terminal fragment about 160 nucleotides long and that this fragment is virtually as efficient in accepting tyrosine as the intact BMV RNAs (18). The present paper reports the complete nucleotide sequence of the fragment derived from RNA 4 and gives data indicating that the fragments from the other three RNAs have nearly the same sequence. They are each 161 bases long, and we designate them O161 of BMV RNAs 1, 2, 3, and 4.

MATERIALS AND METHODS

Materials. T1, T2, and U2 RNases were obtained from Calbiochem. Pancreatic ribonuclease, snake venom phosphodiesterase, and bacterial alkaline phosphatase were obtained from Worthington Biochemical Corp. Nuclease P1 was obtained from Yamasa Shoyu Co., Ltd. (Tokyo, Japan). T4 polyribonucleotide kinase was obtained from P-L Biochemicals. [γ-32P]ATP, at a specific activity of 1500-2000 Ci/mmol, was obtained from Amersham/Searle Corp. Nuclease S1, purified by the method of Vogt (19), was a gift from James E. Dahlberg of the University of Wisconsin. Thin-layer plates (CEL 300 DEAE or CEL 300 DEAE/HR-2/15) were purchased from Macherey-Nagel and Co.

Preparation of BMV RNA and Fragment O161. The procedures for growth and radioactive labeling of BMV and isolation and fractionation of its RNAs have been described (20-22). Partial digestion of the individual BMV RNAs with RNase T1 and subsequent isolation of O161 by electrophoresis on polyacrylamide gels have been described (18). Except where explicitly noted, all our descriptions refer to O161 isolated separately from each of the four RNAs.

Sequence Analyses. Complete digestion of O161 with either T1 RNase or pancreatic ribonuclease, separation of the resulting oligonucleotides by two-dimensional electrophoresis or by electrophoresis-homochromatography, and further analysis of these oligonucleotides were according to the methods of Sanger and his colleagues (23, 24).

The larger oligonucleotides, especially those rich in pyrimidines, were also analyzed by the wandering spot procedure as described by Silberklang et al. (15). Oligonucleotides were labeled at their 5' ends with 32P by use of [γ-32P]ATP and polynucleotide kinase according to the procedures of Simsek et al. (25).

T1 and pancreatic ribonuclease oligonucleotide catalogs were obtained for a variety of large and small pieces of O161, especially those having overlapping sequences. To obtain large

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: BMV, brome mosaic virus.
Fig. 1. Autoradiograms of Ω161 digests. All oligonucleotides were analyzed by pancreatic, T1, U2, and snake venom (preceded by alkaline phosphatase) ribonucleases and some also by the wandering spot technique. (Left) Two-dimensional fractionalization of a complete Ω161 RNase digest of Ω161. The first dimension was electrophoresis on cellulose acetate at pH 3.5; the second dimension was homochromatography on a DEAE-cellulose (DEAE/HR 2/15) thin-layer plate. The dotted circle marked B indicates the position of the blue dye. Note that the oligonucleotide A-C-C-AOH (spot T8) is readily identifiable from its unusual position on the fingerprint. The spots T1–T26 and their relative molar yield are: T1, Gp, 8.4; T2, C-Gp, 1.2; T3, A-Gp, 3.8; T4, C-A-Gp, 0.8; T5, A-A-Gp, 0.9; T6, C-A-A-Gp, 1.1; T7, A-C-A-C-Gp, 0.7; T8, A-C-C-AOH, 0.8; T9, U-U-Gp, 4.0; T10, C-A-U-Gp, 1.0; T11, U-A-C-C-Gp, 0.8; T12, U-A-C-A-Gp, plus T13, C-A-U-A-Gp, 2.0; T14, A-A-U-Gp, 1.0; T15, U-U-U-Gp, 1.1; T16, C-U-U-Gp, plus T17, U-U-C-U-Gp, 2.0; T18, U-C-U-A-Gp, 1.2; T19, A-A-A-A-A-C-A-U-Gp, 0.9; T20, A-C-C-C-U-U-C-U-U-Gp, 0.9; T21, A-C-C-C-U-U-C-U-U-Gp; 1.0; T22, U-A-A-A-U-C-U-C-U-U-A-A-A-Gp, 1.1; T23, C-U-C-U-C-U-U-Gp, 0.9; T24, C-C-U-U-U-Gp, 1.1; T25, U-C-U-U-U-Gp, 0.8; and T26, U-U-A-C-U-C-U-U-U-Gp, 1.0.


overlapping pieces, Ω161 was digested with 20 ng of pancreatic ribonuclease or 400 ng of T1 RNase per 200 μg of RNA. The incubation was at 0° for 20 min. To obtain comparatively shorter pieces (10–30 nucleotides), Ω161 was digested with T1 or pancreatic ribonuclease at an enzyme to substrate ratio of 1:100. The incubation was at 4° for 10 min. The Ω161 fragment was tested for the presence of abnormal bases by two-dimensional thin-layer chromatography on cellulose as described by Nishimura (26). Conditions for digesting Ω161 with nuclease S1 were similar to those of Rushisky and Mozesko (27).

RESULTS

Limited digestion of BMV RNA 4 with T1 RNase and subsequent fractionalization by electrophoresis on polyacrylamide gels produces one major band and many minor bands corresponding to fragments of various lengths (18). The major band fragment, designated Ω161, is the only one produced quantitatively, and it can be obtained readily in pure form. On complete digestion with RNase T1, this fragment gives rise to the 3′-terminal oligonucleotide A-C-C-AOH, indicating that Ω161 is cleaved from the 3′ end of RNA 4.

Sequence Analysis of RNase T1 and Pancreatic Ribonuclease Oligonucleotides of RNA 4 Ω161. Fragment Ω161 of BMV RNA 4, uniformly labeled with 32P (specific activity 109 dpm/mg of RNA), was digested to completion with RNase T1 and the products were separated and characterized. A two-dimensional separation by cellulose acetate electrophoresis and DEAE-cellulose thin-layer homochromatography is shown in Fig. 1 left. All products were well resolved except the isomers T12 and T13 and T16 and T17. Separation of similar quality is obtainable by two-dimensional electrophoresis (18). Fig. 1 right illustrates complete separation by two-dimensional electrophoresis of the products of digestion of Ω161 by pancreatic ribonuclease. The sequences of most of the pancreatic ribonuclease and of the short T1 RNase oligomers were determined by analysis with the complementary RNase, T2 RNase, snake venom phosphodiesterase, and U2 RNase. The oligomers for which such analyses were not definitive were examined, in addition, by the wandering spot procedure; these included T17–T26 and P15, P17, P18, and P21. Some typical wandering spot analyses, those of oligomers T21, T24, T26, and P18, are shown in Fig. 2. Fragment Ω161 of BMV RNA 4 was digested with a mixture of ribonucleases and the products were analyzed for the presence of unusual bases. No spots were detected other than those of the common nucleotides A, C, G, and U.

Ordering of Oligonucleotides and the Sequence of RNA 4 Ω161. Fragment Ω161 of BMV RNA 4 was partially digested with T1 or pancreatic ribonuclease and the pieces were separated by polyacrylamide gel electrophoresis or by cellulose acetate electrophoresis and homochromatography. The purified products were analyzed as had been Ω161 itself. Enough pieces of different chain lengths and with overlapping sequences were obtained to permit ordering of all T1 and pancreatic oligonucleotides except those between residues 70 and 74. We could not distinguish between the sequences C-A-U-G-G-G-C-U-U-G-C-A-U-A-Gp and C-A-U-G-G-C-U-U-G-C-A-U-A-Gp, since both sequences gave rise to the same products after a variety of partial digestions with T1 and pancreatic ribonuclease. This ambiguity was resolved by 32P end-labeling of the T1 partial product corresponding to residues 69–74, followed by wandering spot analysis as shown in Fig. 3.

Sequences of Ω161 from BMV RNAs 1, 2, and 3. The T1 and pancreatic ribonuclease maps for Ω161 from RNAs 1, 2, and 3 were identical to those of RNA 4 except as follows: (a) In Ω161 from RNAs 1 and 2, spot 11 (U-A-C-C-Gp) was missing and was replaced by spots corresponding to the oligomers U-Gp and U-C-Gp in RNA 1 and U-Gp and C-C-Gp in RNA 2. (ii) A-Cp (spot P3) was reduced in amount while G-Up (spot P9) was increased for RNA 1 and G-Cp (spot P4) was increased for RNA 2. Other possible differences are not precluded, e.g., sequence isomers that migrate identically both in electrophoresis and homochromatography. Because extensive differences of this kind are unlikely and would, furthermore, be unlikely to have a bearing on our conclusions, we chose, for the present, to ignore these possibilities.
Fig. 2. Autoradiogram of partial snake venom phosphodiesterase digests (Upper left and right) and partial nuclease P1 digests (Lower left and right) of 5'-32P-labeled spots T21 (Upper left), T24 (Upper right), T26 (Lower left), and P18 (Lower right) of Fig. 1. Complete T1 and pancreatic ribonuclease digests of Ω161 were labeled with [γ-32P]ATP using polynucleotide kinase and the products were separated. 5'-32P-Labeled spots were then eluted, partially digested with the above enzymes, and analyzed. The first dimension was electrophoresis on cellulose acetate, pH 3.5; the second dimension was homochromatography on thin-layer plates made of either pure DEAE-cellulose (CEL 300 DEAE; Upper left and right) or a mixture of DEAE-cellulose and cellulose in the ratio 2:15 (CEL 300 DEAE/HR 2/15; Lower left and right). The homochromatography mixture used was a 3% solution of yeast RNA in 7 M urea that had been hydrolyzed for 45 min, prepared according to Barrett (24). The mononucleotide of T26 was partially trapped in the paper wick.

In all probability RNA 4 Ω161 and RNA 3 Ω161 are identical, RNA 2 Ω161 differs only in that base 46 is G, and RNA 1 Ω161 differs in that base 46 is G and base 45 is U.

Sequence Similarity to tRNAs. The susceptibility of Ω161 to aminocacylation with tyrosine suggests a structural similarity to tyrosine tRNA. Is this resemblance reflected in the Ω161 sequence? Comparison shows that only a limited correspondence to authentic tyrosine tRNAs exists. The sequence of the seven and eight bases adjacent to the 3' terminus of Ω161 is identical to that of the corresponding bases of Saccharomyces cerevisiae and Torula utilis tyrosine tRNA, respectively (28). Also eleven bases in the sequence of Ω161, residues 14–24, have the same sequences as eleven bases in the anticodon loop region of S. cerevisiae, T. utilis, and Escherichia coli tyrosine tRNA except that in these tRNAs, three of the bases are modified. However, the eleven corresponding bases are not in sequentially similar locations with respect to their 3' ends. Thus,

\[
\begin{align*}
\Omega161 & \quad \text{\text{A-C-U-G-U-A-A \ldots A-U-C-U \ldots A-G-A-G-A-C-C-A}} \\
T. \text{utilis} \text{ tyran} & \quad \text{A-C-U-G-A-A \ldots A-U-A-U \ldots A-G-A-G-A-C-C-A} \\
\end{align*}
\]

Other sequence correspondences to Tyr tRNAs are five bases long or less.

Secondary Structure and Base Pairing. Secondary structure is not obviously revealed by inspection of base sequence. Nevertheless some inferences can be drawn by considering structures that maximize the number of Watson–Crick base pairs. For RNA 4 Ω161, the secondary structure of highest stability according to the rules of Tinoco et al. (29) is shown in Fig. 4. As indicated in that figure, all but one T1 RNase cleavage points of high susceptibility are located in the regions lacking base pairs. The single exception is between the C-A bond in positions 77-78. A number of essentially equivalent base-pairing schemes exist. For example, the sequence A-A-G-A-G in positions 4-9 can pair with U-C-U-C-U in positions 103-108. X-ray crystallography shows that in tRNAs the classical cloverleaf model accurately reflects actual base-pairing (although additional base interactions exist that are not revealed by cloverleaf folding). For Ω161, can a cloverleaf structure be drawn with an accessible A-C-C-A terminus and a tyrosine anticodon centered on an “anticodon” loop? Both features are believed to be involved in recognition by the charging enzyme (30). With some alternative folding and with elimination of some marginal base pairs the secondary structure of Fig. 4 can be converted to the more cloverleaf-like structure shown in Fig. 5. However, we are unable to construct a secondary structure of high stability that provides an “anticodon” loop centered on Ω161 bases 14-24.

It is possible, of course, that tertiary interactions, unrevealed by cloverleaf base pairs, determine the charging enzyme recognition features. Regardless of detailed knowledge of tertiary structure, it is to be expected that an enzyme recognition site would be near the surface of a substrate molecule and thus,
might be especially vulnerable to nuclease attack. The U-A bond in the "anticodon" region of the structure in Fig. 5 (positions 66-65) is the most susceptible pancreatic ribonuclease point in Ω161. With an enzyme to substrate ratio of 1:10,000, it was possible to obtain a break only at this point in Ω161, separating the molecule into two parts. The isolated fragments were no longer chargeable. However, a molecule with a hidden break at this point, with the halves still noncovalently bound, was chargeable (M. Bastin and P. Kaesberg, unpublished observations), indicating that the structure on both sides of this anticodon-like feature, although not the integrity of the anticodon itself, is necessary. This is similar to the situation in some tRNAs in which a hidden break in the anticodon loop does not preclude aminoacylation (30).

Fig. 5. Slightly modified secondary structure for RNA 4 Ω161 drawn to illustrate a similarity to the cloverleaf structure of tRNA.

DISCUSSION

Earlier publications and this study have shown that the four BMV RNAs can be enzymatically aminoacylated with tyrosine...
in a manner similar to that of tyrosine tRNA, that a highly susceptible ribonuclease T1 cleavage site exists 161 bases from their 3' terminus, and that in each case the 3'-cleavage product (Q161) can be aminocacylated. Moreover, Q161 is relatively resistant to nuclease digestion; it is obtained in almost quantitative yield over a wide range of T1 concentrations. The Q161 molecules from each of the four RNAs have nearly the same nucleotide sequence. These facts suggest that the 3' end of the BMV RNAs has a structural resemblance to tyrosine tRNA, that it is a tightly folded structure substantially retaining its configuration after cleavage, and (from its sequence conservation) that it plays an important role in the life cycle of the virus.

It was thus gratifying, at least for a short time, that the two longest sequence identities among S. cerevisiae, T. utilis, and E. coli tyrosine tRNA, namely, their aminating terminus and (ignoring base modifications) their anticodon loop region, existed also in Q161. However, these two regions of sequence were separated from each other by 28 bases in the tRNAs but by only four in Q161 and, equally perversely, no manner of stable base-pairing could produce a structure resembling a tRNA anticodon loop and stem. However, folding of Q161 into a cloverleaf-like structure provides a stem and loop structure in which the anticodon AUA (bases 65-67), rather than AUG (bases 19-21) is centered on a loop, as shown in Fig. 5. Moreover, S1 nuclease cleaves preferentially at that site just as it does at the anticodon of tRNA. Thus there is an evident region of resemblance of tertiary structure of Q161 to tRNA but it does not include the 11-base-long region of sequence identity. The significance of the sequence identity is entirely unexplained. It may indicate a sequence recognized by the aminocacylating enzyme, an evolutionary vestige, or coincidence. The problem of recognition of tRNA sites by their aminocacylating enzymes is a complex one and after a variety of studies over a period of many years is only partially solved. (For a comprehensive review, see ref. 30.) The corresponding problem with chargeable viral messenger RNAs is equally formidable and may not be worth pursuing until more is known about the functional significance of the charging.

Our sequence data for Q161 do not provide immediate insight regarding the function of the noncoding region at the 3' end of the BMV RNAs. Whatever that function may be, it has resulted in an evolutionary constraint that provides nearly identical Q161 sequences for the four BMV RNAs even though the proteins encoded in these RNAs are different.

The Q161 regions of the BMV RNAs are not necessarily involved in translation and amino acid transfer even though this is an important function of tRNA. Indeed, there is suggestive evidence to the contrary. BMV RNA that has been chemically modified to preclude acceptance of tyrosine is still fully capable of serving as a messenger in vitro (16). Certainly, pertinent in vitro studies are needed.

Possibly the 3' ends play an important role in initiation of viral assembly.

Possibly the structure at the BMV RNA 3' ends is an important feature of RNA replication. Elongation factors whose normal function is in translation are needed for the replication of the RNA of phage Qβ (32). Tryptophan tRNA serves as a primer in the synthesis of Rous sarcoma viral RNA (33). Perhaps also, with viruses such as BMV, structures nonmimetically associated with translation have a role in replication.

We thank Christopher Sempus and Chris Saris for technical assistance, Dr. G. Altman for computer analysis, and Drs. J. E. Dahlberg and D. Zimmerm for helpful discussions. This research was supported by Grant CA 15613 and Training Grant T32 CA00075 from the National Cancer Institute, Grant AI 01466 and Research Career Award AI 21942 from the National Institute of Allergy and Infectious Diseases, and by Contract 1635 from the Biological Division of the Energy Research and Development Administration.