

EXPERIMENTAL

N.m.r. spectra were normally recorded on 100 MHz (Perkin Elmer R14 and Varian XL100) or 270 MHz (Jeol GX270) spectrophotometers with  $(\text{CD}_3)_2\text{SO}$  as the solvent unless otherwise stated.  $^{31}\text{P}$  N.m.r. spectra were recorded using a Bruker WH400 spectrophotometer operating at 162 MHz. U.v. spectra were measured in ethanol on a Perkin Elmer 552 spectrophotometer. Column chromatography was carried out on silica gel, Kieselgel 60 type 7734, 0.063-0.200 mm, 70-230 mesh ASTM (E. Merck A.G., Darmstadt, W. Germany). All experiments were carried out under scrupulously dry conditions unless otherwise stated and all evaporations of solvents were carried out under reduced pressure. Phosphodiesterase I (EC 3.1.4.1 from *Crotalus adamanteus* venom) was purchased from Worthington, phosphodiesterase II (EC 3.1.16.1 from bovine spleen) was purchased from Pharmacia.

2',3'-Di-O-acetyl-2',3'-secouridine (4). To a solution of 5'-O-trityl-2',3'-secouridine<sup>4</sup> (2) (2.17 g) in pyridine (8 ml) there was added acetic anhydride (1 ml) and the solution kept at room temperature ( $14^\circ$ ) for 2h. The solvent was removed by evaporation and the resultant crude product (3) was dissolved in 80% acetic acid (70 ml) and heated at  $100^\circ$  for 1h. The solvent was again removed by evaporation and the resulting oil was co-evaporated several times with ethanol to give a product which could be purified by column chromatography using chloroform-ethanol (95:5) as eluant. The product (1.38 g, 95%) was obtained as an oil, u.v.  $\lambda_{\text{max}}$  260 nm ( $\epsilon$ , 10,938) (Found: C, 47.1; H, 5.5; N, 8.7.  $\text{C}_{13}\text{H}_{18}\text{N}_2\text{O}_8$  requires C, 47.2; H, 5.45; N, 8.48%); n.m.r.  $\delta$  2.0 and 2.08 (6H, 2s, 2 x  $\text{COCH}_3$ ), 3.35 (2H, d, H-2'), 3.65 (1H, m, H-4'), 4.20 (4H, m, H-5' and H-3'), 4.80 (1H, t, 5'-OH), 5.65 (1H, d, H-5), 6.0 (1H, t, H-1'), 7.70 (1H, d, H-6), 11.35 (1H, bs, NH).

2',3'-Secouridine 5'-p-nitrophenylphosphate (ammonium salt) (6). To a solution of p-nitrophenyl phosphorodichloridate (2.85 g) in dioxan (25 ml) containing pyridine (1.1 ml) there was added a solution of 2',3'-di-O-acetyl-2',3'-secouridine (1.15 g) in dioxan (10 ml) with stirring over a period of 20 min at  $0^\circ$  and the solution kept at  $4^\circ$  for 14h. Aqueous pyridine (1.3 ml pyridine, 7 ml water) was then added, the solvent removed by evaporation and the resultant oil dissolved in ammonia (s.g. 0.880; 40 ml) and stirred overnight at room temperature. The solvent was again removed by evaporation, the residue dissolved in  $\text{H}_2\text{O}$  (300 ml), filtered and 150 ml applied to a diethylaminoethyl cellulose column (DE-52, 2.5 x 60 cm) and eluted with a linear gradient of 0.005-0.1M triethylammonium bicarbonate (TEAB) in a total volume of 2 l. The column was monitored with a u.v. spectrophotometer, the

relevant fractions combined and evaporated repeatedly to dryness to give the required compound as the triethylammonium salt (0.76 g, 83%). This was further purified on a charcoal column (Activated Charcoal for chromatography, Wako Junyaku Co. Ltd., Osaka, Japan, 3 x 24 cm) which was first washed with H<sub>2</sub>O followed by ethanol/ammonia (s.g. 0.880)/H<sub>2</sub>O, 50:2:48 to give the desired product as the ammonium salt as a crystalline solid (0.67 g, 98%), u.v.  $\lambda_{\max}$  273 nm ( $\epsilon$ , 19,970) (Found: C, 36.7; H, 4.9; N 14.9. C<sub>15</sub>H<sub>21</sub>N<sub>4</sub>O<sub>11</sub>P requires C, 38.8; H, 4.55; N, 12.06%); n.m.r. (D<sub>2</sub>O)  $\delta$  3.72 (4H, m, H-5', H-2'), 3.78 (1H, m, H-4'), 4.0 (2H, m, H-3'), 5.92 (2H, m, H-1' and H-5), 7.25 (2H, d, *o*-phenyl), 7.66 (1H, d, H-6), 8.24 (2H, d, *m*-phenyl).

3'-O-Mesyl-2',3'-secouridine (9). A solution of O<sup>2</sup>,2'-anhydro-3'-O-mesyl-5'-O-trityl-2',3'-secouridine<sup>2</sup> (8) (12.4 g) in 80% acetic acid (120 ml) was heated at 100° for 2h. The solvent was removed to give an oil which was purified by column chromatography using methanol-chloroform (2:10) as eluant. The product (4.32 g, 59%) was obtained as a white foam, u.v.  $\lambda_{\max}$  ( $\epsilon$ , 9,800) (Found: C, 37.0; H, 5.2; N, 8.0. C<sub>10</sub>H<sub>16</sub>N<sub>2</sub>O<sub>8</sub>S requires C, 37.03; H, 4.93; N, 8.64%); n.m.r.  $\delta$  3.20 (3H, s, mesyl), 3.31-3.80 (5H, m, H-5', H-2' and H-4'), 4.83 and 5.13 (2H, bs, 2'-OH and 5'-OH), 5.64 (1H, d, H-5), 5.81 (1H, t, H-1'), 7.65 (1H, d, H-6), 11.15 (1H, bs, NH).

3'-O-Mesyl-2',5'-di-O-p-monomethoxytrityl-2',3'-secouridine (10). To a solution of compound 9 (4.24 g) in pyridine (40 ml) there was added p-monomethoxytrityl chloride (8.87 g) and the solution stirred at room temperature overnight. Water was then added, the solvent removed by evaporation and traces of pyridine removed by co-evaporation with toluene. The resulting oil was fractionated by column chromatography using toluene-acetone (8:2) as eluant to give the product as a white foam (9.31 g, 82%), u.v.  $\lambda_{\max}$  260 nm ( $\epsilon$ , 15,650) (Found: C, 68.9; H, 5.7; N, 3.0. C<sub>50</sub>H<sub>48</sub>N<sub>2</sub>O<sub>16</sub>S requires C, 69.1; H, 5.56; N, 3.22%); n.m.r. (CDCl<sub>3</sub>)  $\delta$  2.85 (3H, s, mesyl), 3.25 (2H, m, H-5'), 3.45 (1H, m, H-4'), 3.65-3.87 (8H, m, H-2' and 2 x OCH<sub>3</sub>), 4.36 (2H, m, H-3'), 5.53 (1H, d, H-5), 6.05 (1H, t, H-1'), 6.82 (4H, d, 2 x phenyl *o*-to -OCH<sub>3</sub>), 7.15-7.50 (25H, m, phenyl and H-6), 8.72 (1H, bs, NH).

3'-O-Acetyl-2',5'-di-O-p-monomethoxytrityl-2',3'-secouridine (11). To a solution of compound 10 (9.2 g) in DMF (50 ml) was added potassium acetate (2.59 g) and the suspension stirred while heating at 100° for 2h when the suspension cleared. Saturated sodium chloride solution was then added and the mixture extracted with ethyl acetate (3 x 80 ml). The organic layers were combined, washed with H<sub>2</sub>O, dried and evaporated to dryness to yield an oil which was purified by column chromatography using toluene-acetone (8:2) as

eluant. The product (6.95 g, 79%) was obtained as a foam, u.v.  $\lambda_{\max}$  260 nm ( $\epsilon$ , 14,800) (Found: C, 73.5; H, 5.6; N, 3.2.  $C_{51}H_{48}N_2O_9$  requires C, 73.54; H, 5.80; N, 3.38%); n.m.r.  $\delta$  1.95 (3H, s, acetyl), 3.2 (3H, m, H-2' and H-5'), 3.4 (1H, m, H-5'), 3.77 (7H, m, H-4' and 2 x  $OCH_3$ ), 4.12 (1H, m, H-3'), 4.32 (1H, dd, H-3'), 5.52 (1H, d, H-5), 6.10 (1H, t, H-1'), 7.76 (4H, d, 2 x phenyl o- to  $-OCH_3$ ), 7.18 (24H, m, phenyl), 7.32 (1H, d, H-6), 8.98 (1H, bs, NH). 2',5'-Di-O-p-monomethoxytrityl-2',3'-secouridine (12). To a solution of compound 11 (6.82 g) in a mixture of dioxan (30 ml) and methanol (100 ml) was added ammonia gas until saturated with cooling. The solution was left at room temperature for 48h, the solvent removed by evaporation and the residue left, purified by column chromatography using toluene-acetone (8:2) as eluant. The product (4.91 g, 76%) was obtained as a foam, u.v.  $\lambda_{\max}$  265 nm ( $\epsilon$ , 14,400) (Found: C, 74.2; H, 5.6; N, 3.7.  $C_{49}H_{46}N_2O_8$  requires C, 74.41; H, 5.86; N, 3.54%); n.m.r.  $\delta$  3.10 (2H, m, H-2'), 3.35 (2H, m, H-5'), 3.44 (2H, m, H-3'), 3.55 (1H, m, H-4'), 3.72 (6H, s, 2 x  $OCH_3$ ), 4.80 (1H, t, 3'-OH), 5.52 (1H, d, H-5), 6.04 (1H, t, H-1'), 6.85 (4H, dd, 2 x phenyl o- to  $-OCH_3$ ), 72.5 (24H, m, phenyl), 7.5 (1H, d, H-6), 11.45 (1H, bs, NH). 2',3'-Secouridine 3'-p-nitrophenylphosphate (ammonium salt) (14). To a solution of p-nitrophenylphosphorodichloridate (0.566 g) in dioxan (2.0 ml) and pyridine (0.216 ml) was added dropwise a solution of compound 12 (0.547 g) in dioxan (3 ml) with cooling to 0° and stirring. The reaction mixture was stirred at room temperature for 2h and then left at 4° for 12h. Aqueous pyridine (0.25 ml pyridine, 1.2 ml  $H_2O$ ) was added rapidly with stirring, the solvent removed by evaporation to give the crude product (13). This compound was dissolved in a mixture of chloroform and  $H_2O$ , the organic layer separated, dried, taken to dryness, the residue dissolved in 80% acetic acid (25 ml) and the solution kept at room temperature for 18h. The solvent was removed by evaporation and the residue co-evaporated with water to remove traces of acetic acid. The resulting solid was partitioned between ether and water and the aqueous layer (250 ml) applied to a DEAE-column (DE-52, 3.0 x 35 cm), eluted with 0.005M TEAB (1.5 l) and then with a linear gradient of 0.005-0.075M TEAB in a total volume of 2 l. The column was monitored with a u.v. spectrophotometer, the relevant fractions combined and evaporated repeatedly to dryness to give the required compound as the triethylammonium salt (0.226 g, 60%). This was further purified on a charcoal column and the desired product isolated as previously described for compound (6) as the ammonium salt, u.v.  $\lambda_{\max}$  267 nm ( $\epsilon$ , 19,500); n.m.r. ( $D_2O$ )  $\delta$  3.80 (2H, m, H-5), 3.95 (3H, m, H-2', H-4'), 4.25 (2H, m, H-3'), 6.06 (1H, d, H-5), 6.12 (1H, t,

H-1'), 7.56 (2H, d, o-phenyl), 7.96 (1H, d, H-6), 8.48 (2H, d, m-phenyl).  
H.p.l.c. analysis of compounds 6 and 14 showed no significant impurities and no cross-contamination.

ACKNOWLEDGEMENTS

The authors thank the SERC for a Fellowship (to S.N.) and for supporting this work.

REFERENCES

1. This paper is dedicated to Professor Morio Ikehara on the occasion of his retirement from Osaka University in March 1986.
2. Jones, A.S., Walker, R.T., Wyatt, P.G., Balzarini, J. and De Clercq, E. (1985) *J. Chem. Res.* 336-338.
3. Jones, A.S., McLean, M.J., Tanaka, H., Walker, R.T., Balzarini, J. and De Clercq, E. (1985) *Tetrahedron* 41, 5965-5972.
4. Howarth, O., Jones, A.S., Walker, R.T. and Wyatt, P.G. (1984) *J. Chem. Soc. Perkin Trans. II*, 261-265.

---

**The nucleotide sequence of tobacco vein mottling virus RNA**

---

Leslie L.Domier<sup>1</sup>, Kathleen M.Franklin<sup>+</sup>, Muhammed Shahabuddin, Gary M.Hellmann, Jean H.Overmeyer, Shivanand T.Hiremath, Martin F.E.Siaw<sup>2</sup>, George P.Lomonosoff<sup>3+</sup>, John G.Shaw<sup>+</sup> and Robert E.Rhoads<sup>4</sup>

---

Departments of Biochemistry and <sup>+</sup>Plant Pathology, University of Kentucky, Lexington, KY 40536, USA

---

Received 14 April 1986; Revised and Accepted 11 June 1986

---

**ABSTRACT**

The nucleotide sequence of the RNA of tobacco vein mottling virus, a member of the potyvirus group, was determined. The RNA was found to be 9471 residues in length, excluding a 3'-terminal poly(A) tail. The first three AUG codons from the 5'-terminus were followed by in-frame termination codons. The fourth, at position 206, was the beginning of an open reading frame of 9015 residues which could encode a polyprotein of 340 kDa. No other long open reading frames were present in the sequence or its complement. This AUG was present in the sequence AGGCCAUG, which is similar to the consensus initiation sequence shared by most eukaryotic mRNAs. The chemically-determined amino acid compositions of the helper component and coat proteins were similar to those predicted from the nucleotide sequence. Amino acid sequencing of coat protein from which an amino-terminal peptide had been removed allowed exact location of the coat protein cistron. A consensus sequence of V-(R or K)-F-Q was found on the N-terminal sides of proposed cleavage sites for proteolytic processing of the polyprotein.

**INTRODUCTION**

The potyviruses are a large group of flexuous rod-shaped plant viruses which contain 10-kb, monopartite, (+)-sense, single-stranded RNA genomes. Several investigations (1-10) have established the approximate positions in potyviral genomes of cistrons encoding five proteins: coat protein (CP), cylindrical inclusion protein (CI), helper component protein (HC), and two nuclear inclusion proteins (designated NI<sub>a</sub> and NI<sub>b</sub> in this report). In addition, a sixth polypeptide, the genome-linked protein (VPg; 11,12) is likely to be encoded by the viral RNA. Evidence has been presented for post-translational proteolytic processing of polyproteins as a mechanism of potyviral genome expression (5,6,9,10,13-15).

In order to more fully understand the structure and function of the potyviral genome, we have determined the nucleotide sequence of the RNA of one member of this group, tobacco vein mottling virus (TVMV).

## MATERIALS AND METHODS

### Materials

Avian myeloblastosis virus reverse transcriptase was obtained from Life Sciences, Inc., St. Petersburg, FL. Restriction endonucleases and *Escherichia coli* DNA polymerase I were obtained from New England Biolabs, Beverly, MA. T4 DNA ligase was prepared by the method of Davis *et al.* (17). TMV RNA was purified as previously described (4).

### Construction of Recombinant Plasmids

Synthesis of cDNA copies of the 5'-terminus of TMV RNA was primed with a synthetic tetradecadeoxynucleotide complementary to residues 1131 through 1144. The second-strand was synthesized using *E. coli* DNA polymerase I and RNase H (18). Residues of dG were added using terminal deoxyribonucleotidyl transferase. The ds cDNA was annealed with pUC18 which had been cleaved with *Pst*I and to which a dC tail had been added, and the product was used to transform *E. coli* strain JM83 (19).

Plasmids containing 3'-terminal sequences were constructed using cDNA synthesized by oligo(dT) priming. Following second-strand synthesis with *E. coli* DNA polymerase I, the cDNA was treated with S1 nuclease, a dC tail added, and the cDNA inserted into pB322 which had been cleaved with *Pst*I and to which a dG tail had been added (20).

### Nucleotide Sequence Analysis

TMV cDNA inserts from five previously described recombinant plasmids pTV-H1, pTV-H2, pTV-H3, pTV-H4 and pTV-P1 (21), as well as from pTV-H7 and pTV-1L (see RESULTS AND DISCUSSION), were inserted into M13mp19 replicative form DNA and prepared for sequencing (22,23). The nucleotide sequences of the inserts were determined by the chain termination method (24) as modified by Biggin *et al.* (25). In some cases, reverse transcriptase was substituted for the large fragment of *E. coli* DNA polymerase I. Portions of the single-stranded DNA inserts of the M13 subclones were sequentially deleted (26) to permit further sequence analysis.

Direct RNA (27) and cDNA sequence analysis were performed by a modified chain termination method using synthetic oligodeoxynucleotide primers and reverse transcriptase. First strand cDNA was synthesized as previously described (21). Oligodeoxynucleotide primers used in both RNA and DNA sequence analysis were prepared using an Applied Biosystems Model 380A DNA synthesizer.

Sequences were compiled and analyzed using an IBM PC computer and software developed by Queen and Korn (28).

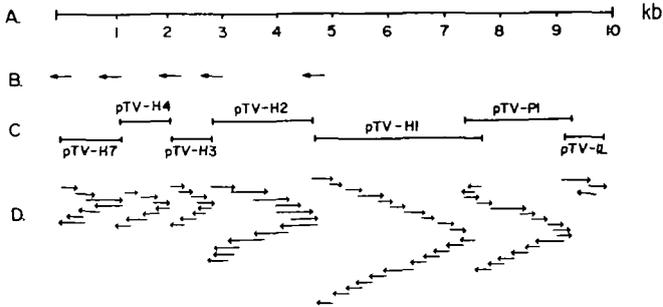


Fig 1. Sequencing strategy for TVMV RNA. A, nucleotide residue number (in kb) beginning with the 5'-terminus of the viral RNA. B, sequence derived from oligodeoxynucleotide primers used to directly determine RNA sequence. The direction of DNA chain elongation is indicated by the arrows. C, TVMV cDNA inserts cloned into recombinant plasmids. The nucleotide locations of the inserts are: pTV-H7, 35 - 1144; TV-H4, 1105 - 2093; pTV-H3, 2094 - 2851; pTV-H2, 2852 - 4573; pTV-H1, 4574 - 7586; pTV-P1, 7254 - 8975; pTV-L, 8813 - 9471. D, single-stranded DNA templates obtained by subcloning and sequentially deleting portions of the inserts of recombinant plasmids in M13 vectors. The arrows show the location, direction, and length of nucleotide sequence obtained from each template.

### Analysis of TVMV Proteins

TVMV CP and HC were purified as previously described (4,29). Amino acid compositions were determined by the method of Hirs *et al.* (30) using a Beckman System 6300 amino acid analyzer.

Purified TVMV was treated with trypsin as described by Allison *et al.* (15) for tobacco etch virus (TEV). The amino acid sequence of the resulting trypsin-treated CP was determined by the trifluoroacetic acid conversion method (31) using an Applied Biosystems gas-phase sequencer.

## RESULTS AND DISCUSSION

### Characterization of Plasmids Containing 5'- and 3'-terminal Sequences of TVMV RNA

Since the previously isolated recombinant plasmids did not contain cDNA copies of the 5'- or 3'-termini of TVMV RNA (12,21), it was necessary to construct plasmids which represented these regions. DNA complementary to the 5'-terminus was generated by priming synthesis with an oligodeoxynucleotide complementary to sequences in pTV-H4 nearest the 5'-terminus of the RNA. The plasmid containing the largest 5'-terminal insert, designated pTV-H7, included the priming site in pTV-H4 for cDNA synthesis and an additional 1070 nucleotides toward the 5'-terminus of the viral genome (Fig. 1C).

The sequence of the 5'-terminal region of TVMV RNA which was not contained within pTV-H7 was determined by direct RNA sequencing using reverse transcriptase and a synthetic oligodeoxynucleotide primer (Fig. 1B). This provided evidence for only 35 additional nucleotides beyond the terminus of pTV-H7, suggesting that the 5'-terminus of the RNA had been reached. Verification of the 5'-terminal pentanucleotide sequence was made by wandering spot analysis of RNase T1-digested, [<sup>32</sup>P]pCp-labeled, VPg-linked RNA (Shahabuddin, Shaw and Rhoads, manuscript in preparation).

The insert of one recombinant plasmid generated from the 3'-terminus, pTV-1L, was found to include a terminal poly(A) segment as well as an overlap with pTV-P1 of approximately 200 nucleotides (Fig. 1C). The presence of a poly(A) tail has been reported for TVMV RNA based on the retention of the RNA by oligo(dT)-cellulose and the ability to prime synthesis of apparently full-length cDNA using oligo(dT) (4,5). The presence of poly(A) tracts at the 3'-termini of the RNA of two other potyviruses, TEV (6) and pepper mottle virus (9), have been directly determined. These are evidence that the poly(A) tract found in pTV-1L does, in fact, represent the 3'-terminus of the TVMV RNA.

The sequences of both strands of all recombinant plasmid inserts were determined after subcloning into M13mp19 (Fig. 1D). The sequence of pTV-1L was confirmed by sequence analysis of first strand cDNA using oligodeoxynucleotides to prime DNA synthesis. To verify that the non-overlapping recombinant plasmids (pTV-H1 through -H4) represented contiguous sequences of viral RNA, a series of oligodeoxy- nucleotides was synthesized in order to prime cDNA synthesis beginning 30-40 nucleotides from each of the proposed HindIII junctions (Fig. 1B). This analysis revealed that the plasmid inserts from pTV-H1 through -H4 were contiguous. Analysis of the nucleotide sequence in the overlap region between pTV-H7 and pTV-H4 revealed the presence of two HindIII sites separated by 34 nucleotides. It was found that when pTV-H7 had been cleaved with HindIII and subcloned into M13, only the larger of the resulting cDNA fragments had been isolated. Consequently, an EcoRI fragment of pTV-H7 containing the 34-bp region was subcloned into M13 and its sequence determined.

#### Primary Structure of TVMV RNA

The nucleotide sequence is presented in Fig. 2. The RNA contains 9471 nucleotides excluding the 3'-terminal poly(A). Base composition of the viral RNA sequence revealed a high adenine content (32.0%) followed by uracil (26.1%), guanine (22.9%), and cytosine (19.0%). These compositions are similar to those reported for cowpea mosaic virus (CPMV; 32,33) and human







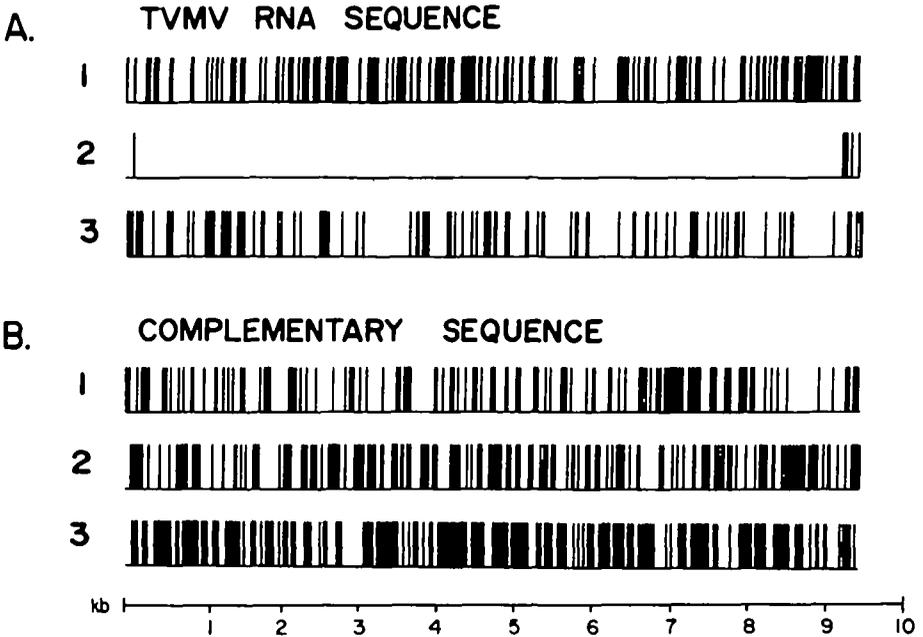


Fig 3. Termination Codons in TVMV RNA. Computer translation of the sequence in Fig. 2 was performed in all three reading frames for both the viral and complementary nucleotide sequence. The numbers on the left indicate the three different reading frames. Each vertical line represents the location of a termination codon in the sequence.

found at positions 3, 153 and 165 and would encode peptides of only 8, 35 and 31 amino acid residues, respectively. The sequence surrounding the AUG codon at position 206, AGGCCAUG, is in reasonable agreement with the consensus ribosomal recognition sequence of eukaryotic mRNAs, CCRCCAUG, noted by Kozak (36). Other potential initiation codons downstream occur in less favorable contexts. This observation, plus its location at the beginning of the long open reading frame, make the AUG at position 206 the best candidate for the initiation codon of the putative polyprotein. This assignment is also in agreement with our previous estimate, based on hybrid-arrested translation of the HC cistron, that the untranslated leader of TVMV RNA is 200 nucleotides in length (35).

Unlike the RNAs of CPMV and the picornaviruses, TVMV RNA contains a potential polyadenylation signal, AAUAAA, in its 3' nontranslated region at position 9377. Polyadenylation signal sequences are usually found within 25 residues of the site of poly(A) addition (16). The AAUAAA sequence of TVMV

Table 1. Comparison of predicted and chemically determined amino acid compositions of TVMV helper component and coat protein

Helper Component			Coat Protein		
Amino acid	Predicted <sup>a</sup>	Chemically Determined <sup>b</sup>	Amino acid	Predicted <sup>a</sup>	Chemically Determined <sup>b</sup>
residues per mole					
Ala	41	38	Ala	24	23
Arg	27	27	Arg	14	12
Asx	63	63	Asx	32	33
Cys	13	ND <sup>c</sup>	Cys	2	ND
Glx	41	47	Glx	28	30
Gly	28	35	Gly	17	21
His	12	9	His	9	8
Ile	28	28	Ile	10	6
Leu	54	58	Leu	18	24
Lys	28	32	Lys	17	17
Met	8	11	Met	13	10
Phe	21	22	Phe	11	9
Pro	17	25	Pro	12	11
Ser	35	28	Ser	13	18
Thr	26	21	Thr	16	14
Trp	3	ND	Trp	3	ND
Tyr	17	16	Tyr	5	5
Val	22	30	Val	21	19

<sup>a</sup>Predicted from the nucleotide sequence.

<sup>b</sup>Calculated by amino acid analysis of the isolated proteins.

<sup>c</sup>Not determined

RNA, however, is found 94 bases from the start of the poly(A). The AAUAAA sequence is also found at seven other positions in the TVMV genome.

#### Secondary Structure in the 5'- and 3'- Non-coding Regions of TVMV RNA

Non-translated regions at the 5'- and 3'-termini of TVMV RNA were analyzed for dyad symmetry using parameters set by Tinoco *et al.* (37). Potential stem-loop structures could be generated in the 5'-non-coding region between nucleotide residues 147 and 171 ( $\Delta G = -6.8$  kcal/mol), and in the 3'-non-coding region, between nucleotide residues 9233 and 9250 ( $\Delta G = -3.8$  kcal/mol). Such secondary structures in DNA and RNA sequences of both prokaryotes and eukaryotes have been reported to be associated with regulation of gene expression (38,39,40). In viruses, these structures have been suggested to be involved in replication, regulation of transcription or virion assembly (41,42,43,44).

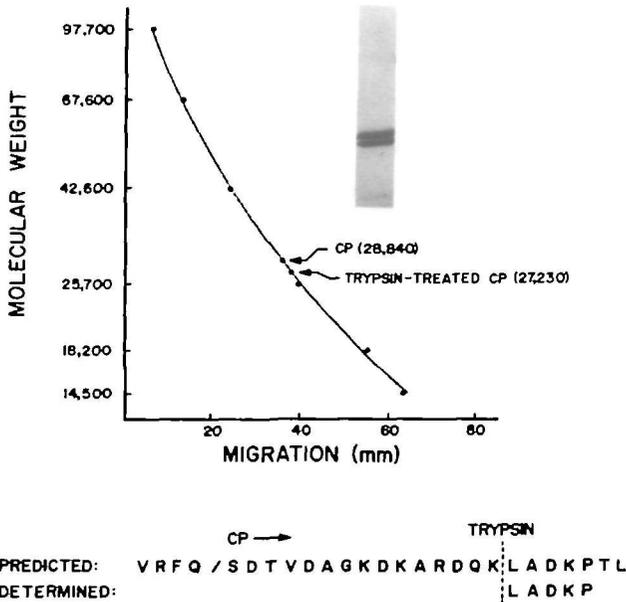


Fig 4. Determination of amino acid sequence of trypsin-treated TVMV coat protein (CP). Purified TVMV was treated with trypsin and analyzed by SDS-PAGE, with protein visualized by Coomassie blue staining (inset). The molecular weight of the resulting CP fragment was estimated by electrophoretic mobility. The amino acid sequence of the polypeptide in the region of the proposed cleavage site (//) separating CP from NI<sub>5</sub>, as predicted from the nucleotide sequence, and that obtained by Edman degradation of the tryptic digest, are shown.

#### Location of CP and HC Cistrons

The locations of a number of TVMV cistrons have been proposed based on hybrid-arrested translation, expression of cDNA fragments in bacterial cells, and immunoprecipitation of *in vitro* translation products with antisera to isolated viral proteins (10,35). Further confirmation for the locations of two of these cistrons is provided by comparison of amino acid compositions predicted from the RNA sequence with those determined chemically with isolated TVMV proteins. The results of such comparisons show reasonable agreement between predicted and analyzed compositions for both CP and HC (Table 1).

Attempts at determining the amino acid sequence of intact TVMV CP by chemical methods were not successful, presumably due to a blocked N-terminus, a situation also encountered with some isolates of TEV (15). Allison *et al.* (15) were able to obtain amino acid sequence data near the N-terminus of the

CP of HAT isolates of TEV, however, after trypsin treatment of intact virions which apparently removed the N-terminal 29 amino acid residues of CP. When purified TVMV was likewise treated with trypsin, a form of CP was obtained having a molecular weight of 27,230, slightly lower than the native molecular weight of 28,840 (Fig. 4). This reduction in molecular weight suggested that approximately 15 amino acid residues had been cleaved from the CP. Chemical sequencing of the tryptic digest yielded an initial sequence of five amino acids which was identical to amino acid residues 16-20 of the predicted CP sequence (amino acid positions 2756-2760 of the polyprotein; Fig. 2). This sequence of five amino acids was found at no other location in the putative polyprotein. Tryptic cleavage after the lysine residue at position 15 of the predicted CP would yield a C-terminal fragment with a molecular weight of 27,900, similar to that estimated by SDS-polyacrylamide gel electrophoresis. Thus, we propose that CP is produced by proteolytic cleavage of the TVMV polyprotein between the glutamyl residue at amino acid position 2740 and the following seryl residue, the same dipeptide apparently cleaved in TEV HAT isolates (15). Comparison of the RNA sequences of TEV and TVMV reveals that the predicted cleavages occur at exactly the same amino acid positions from the C-terminus. It is interesting that the CP of both TVMV and HAT isolates of TEV is apparently produced by cleavage at Gln-Ser and is N-blocked while the CP of NAT isolates of TEV appears after cleavage at Gln-Gly and is not N-blocked (6).

#### Assignment of Viral Polyprotein Cleavage Sites

It has been proposed that mature potyviral proteins are produced from a polyprotein which is proteolytically cleaved (5,6,9,10,13-15). This mechanism is similar to that employed by the picornaviruses (45,47) and CPMV (32,46). The poliovirus polyprotein is cleaved primarily at Gln-Gly sites. It has been proposed that the CPMV polyproteins are cleaved at selected Gln-Gly, Gln-Met and Gln-Ser sites. The positions of all Gln-Ala, Gln-Gly, and Gln-Ser dipeptides in the TVMV polyprotein are shown in Fig. 5. Using the cleavage site predicted for CP (Fig. 4) and the approximate positions of TVMV cistrons (10), it was possible to find Gln-Ala, Gln-Gly and Gln-Ser sites corresponding to the proposed termini of each cistron.

Examination of the amino acid sequence in the vicinity of five of the selected dipeptides revealed a high degree of homology on the N-terminal side of the cleavage sites (Fig. 5 and Table 2). A consensus cleavage sequence of V-(R or K)-F-Q/(G, S or A) emerged, where "/" indicates the point of cleavage. A sixth site, near the proposed N-terminus of the NI<sub>a</sub> cistrons,

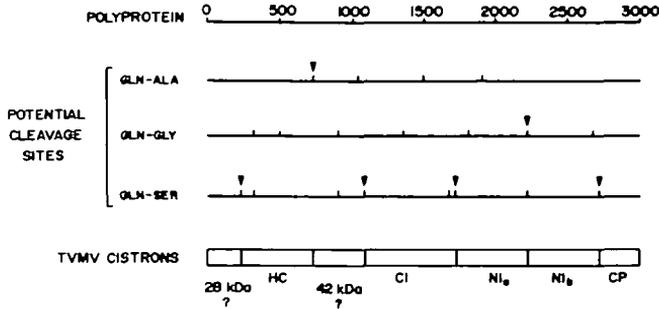


Fig 5. Proposed protease cleavage map of the TVMV polyprotein. Computer analysis was performed to localize the positions (vertical lines) of the dipeptides Gln-Ala, Gln-Gly, and Gln-Ser in the putative TVMV polyprotein sequence (horizontal lines). Those which represent the proposed cleavage sites for various TVMV proteins are indicated by arrows. The amino acid positions proposed to represent the termini of each protein are given in Table 2. The cistron map of TVMV RNA (10) is shown at the bottom.

showed homology to the consensus cleavage site. If utilized, it would remove a polypeptide of 5.5 kDa from the NI<sub>a</sub> protein. The other Gln-X sites not located at the proposed termini of cistrons did not share this homology. Proteolytic cleavage of the TVMV polyprotein at the predicted sites would generate the five known TVMV proteins as well as two other potential polypeptides, a 28-kDa polypeptide located at the N-terminus and a 42-kDa polypeptide located between HC and CI.

The molecular weights of the five polypeptides known to be encoded by TVMV RNA, calculated from these proposed proteolytic cleavage sites, are in reasonable agreement with the molecular weights of the respective TVMV proteins determined by electrophoretic mobility (Table 2).

Table 2. Comparison of experimentally determined molecular weights of TVMV proteins with those predicted from the nucleotide sequence

TVMV Protein	MW X 10 <sup>-3</sup> Measured <sup>a</sup>	Literature References	MW X 10 <sup>-3</sup> Predicted <sup>b</sup>	Amino Acid Positions
HC	53	48	54.0	248-731
CI	70	16	71.1	1113-1747
NI <sub>a</sub>	52	10	53.3	1748-2224
NI <sub>b</sub>	56	10	59.1	2225-2740
CP	29	this report	29.6	2741-3005

<sup>a</sup>Estimated from SDS polyacrylamide gel electrophoresis in the indicated studies.

<sup>b</sup>Predicted from the nucleotide sequence using the cleavage sites shown in Fig. 5.

## ACKNOWLEDGEMENTS

This work was supported by Grant 4E021 from the University of Kentucky Tobacco and Health Research Institute, Grants 83-CRCR-1-1-1258 and 58-7B30-3-538 from the USDA and Grant 563/82 from NATO. The authors are grateful to Dr. Frank Gannon and Dr. Pierre Chambon, University of Louis Pasteur, Strasbourg, France, in whose laboratory this work was initiated.

<sup>1</sup>Supported by a grant from the Brown and Williamson Tobacco Corporation, Louisville, KY.

<sup>2</sup>Current address: Research Institute of Scripps Clinic, 10660 North Torrey Pines Road, La Jolla, CA 92037.

<sup>3</sup>Current address: Department of Virus Research, John Innes Institute, Colney Lane, Norwich NR4 7UH, United Kingdom.

<sup>4</sup>To whom correspondence should be addressed.

## REFERENCES

1. Dougherty, W.G., Hiebert, E. (1980) *Virology* **101**, 466-474.
2. Dougherty, W.G., Hiebert, E. (1980) *Virology* **104**, 174-182.
3. Dougherty, W.G., Hiebert, E. (1980) *Virology* **104**, 183-194.
4. Hellmann, G.M., Shaw, J.G., Lesnaw, J.A., Chu, L.-Y., Pirone, T.P., Rhoads, R.E. (1980) *Virology* **106**, 207-216.
5. Hellmann, G.M., Thornbury, D.W., Hiebert, E., Shaw, J.G., Rhoads, R.E. (1983) *Virology* **124**, 434-444.
6. Allison, R.F., Sorenson, J.C., Kelly, M.E., Armstrong, F.B., Dougherty, W.G. (1985) *Proc. Natl. Acad. Sci USA* **82**, 3969-3972.
7. deMejia, M.V.G., Hiebert, E., Purcifull, D.E., Thornbury, D.W., Pirone, T.P. (1985) *Virology* **142**, 34-43.
8. Nagel, J., Hiebert, E. (1985) *Virology* **143**, 435-441.
9. Dougherty, W.G., Allison, R.F., Parks, T.D., Johnston, R.E., Feild, M.J., Armstrong, F.B. (1985) *Virology* **146**, 282-291.
10. Hellmann, G.M., Hiremath, S.T., Shaw, J.G., Rhoads, R.E. (1986) *Virology*, in press.
11. Hari, V. (1981) *Virology* **112**, 391-399.
12. Staw, M.F.E., Shahabuddin, M., Ballard, S., Shaw, J.G., Rhoads, R.E. (1985) *Virology* **142**, 134-143.
13. Vance, V.B., Beachy, R.N. (1984) *Virology* **132**, 271-281.
14. Yeh, S.D., Gonsalves, D. (1985) *Virology* **143**, 260-271.
15. Allison, R.F., Dougherty, W.G., Parks, T.D., Willis, L., Johnston, R.E., Kelly, M.E., Armstrong, F.B. (1985) *Virology* **147**, 309-316.
16. Berget, S.M. (1984) *Nature* **309**, 179-182.
17. Davis, R.W., Botstein, D., Roth, J.R. (1980) *Advanced Bacterial Genetics*, pp. 196-197. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
18. Gubler, U., Hoffman, B.J. (1983) *Gene* **25**, 263-269.
19. Yanisch-Perron, C., Vieira, J., Messing, J. (1985) *Gene* **33**, 103-119.
20. Efstratiadis, A., Villa-Komaroff L. (1979) in Stelow, J.K. and Hollaender, A. (eds), *Genetic Engineering*, Plenum Press, New York, Vol. 1, p. 15.
21. Hellmann, G.M., Shahabuddin, M., Shaw, J.G., Rhoads, R.E. (1983) *Virology* **128**, 210-220.