Volume 14 Number 13 1986 Nucleic Acids Research

The nucleotide sequence of tobacco vein mottling virus RNA

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

Departments of Biochemistry and +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), cylindirical inclusion protein (CI), helper component protein (HC), and two nuclear inclusion proteins (designated NI_a and NI_b 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). TVMV RNA was purified as previously described (4).

Construction of Recombinant Plasmids

Synthesis of cDNA copies of the 5'-terminus of TVMV 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 PstI 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 PstI and to which a dG tail had been added (20).

Nucleotide Sequence Analysis

TVMV 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).

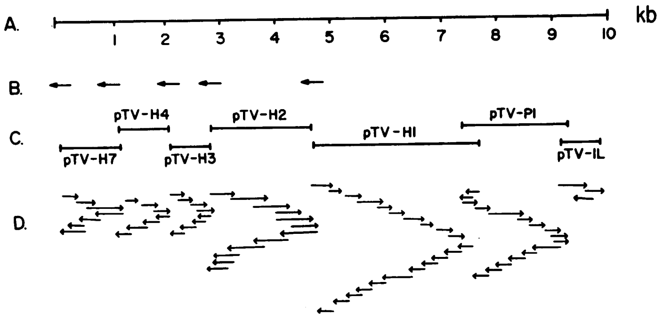

Fig 1. Sequencing strategy for TVMV RNA. A, nucleotide residue number (in kb) beginning with the 5'-terminus of the viral RNA. B, sequence derived from oligodeoxynucleotide primers used to directly determine RNA sequence. The direction of DNA chain elongation is indicated by the arrows. C, TVMV cDNA 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-1L, 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 oligodeoxynucleo- tide 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, [^{32}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

```
AAAUGAAACAAAUCAACACAACAUUAUAACGAACCAGCAAUCUCAAGCAAUCAAGCUAUUCUCAGCAAUUUCAGCAAACACAACUACAGAAAGUAAUUUUUCACUCAAUUAAUUUUCAUU
                            30                            60                            90                           120
                                                                                                               10
                                                                                     M  S  T  I  H  S  A  V  T  A  E  K
AGUUUUUAACUGCGACAAUAGCAGAGAGAUCAAUGGCAGCAACAAUGAUCUUUGGUUCCUUCACUCACGAUCUUUUGGCAAGGCCAUGUCAACCAUUCACUCAGCCGUCACAGCUGAGAA
                           150                           180                           210                           240
                     20                            30                            40                            50
 D  I  F  S  S  I  K  E  R  L  E  R  K  R  H  G  K  I  C  R  M  K  N  G  S  I  Y  I  K  A  A  S  S  T  K  V  E  K  I  N
AGACAUAUUCUCCAGCAUAAAGGAGCGUCUUGAAAGAAAGAGGCAUGGGAAAAUAUGCCGAAUGAAAAACGGCAGCAUAUACAUCAAGGCUGCUUCGUCCACGAAAGUGGAGAAAAUAAA
                           270                           300                           330                           360
                     60                            70                            80                            90
 A  A  A  K  K  L  A  D  D  K  A  A  F  L  K  A  Q  P  T  I  V  D  K  I  I  V  N  E  K  I  Q  V  V  E  A  E  E  V  H  K
CGCAGCAGCCAAGAAGCUAGCCGAUGAUAAAGCAGCUUUCUUGAAGGCUCAACCAACAAUUGUUGACAAAAUCAUUGUCAAUGAGAAGAUACAAGUGGUGGAAGCUGAAGAAGUGCACAA
                           390                           420                           450                           480
                     100                           110                           120                           130
 R  E  D  V  Q  T  V  F  F  K  K  T  K  K  R  A  P  K  L  R  A  T  C  S  S  S  G  L  D  N  L  Y  N  A  V  A  N  I  A  K
GCGAGAGGAUGUGCAAACUGUAUUCUUUAAGAAAACCAAAAAGAGGGCGCCCAAGUUGCGCGCAACUUGUAGCAGCAGUGGUUUAGAUAAUUUGUAUAAUGCAGUGGCUAAUAUAGCCAA
                           510                           540                           570                           600
                     140                           150                           160                           170
 A  S  S  L  R  V  E  V  I  H  K  K  R  V  C  G  E  F  K  Q  T  R  F  G  R  A  L  F  I  D  V  A  H  A  K  G  H  R  R  R
AGCAAGUUCUCUUCGAGUGGAGGUUAUCCACAAAAAGAGGGUUUGUGGAGAGUUCAAGCAAACAAGGUUUGGUAGAGCCUUGUUCAUUGACGUUGCCCACGCAAAAGGCCACAGACGUCG
                           630                           660                           690                           720
                     180                           190                           200                           210
 I  D  C  R  M  H  R  R  E  Q  R  T  M  H  M  F  M  R  K  T  T  K  T  E  V  R  S  K  H  L  R  K  G  D  S  G  I  V  L  L
CAUCGAUUGCAGAAUGCAUAGACGUGAGCAGCGAACUAUGCACAUGUUCAUGCGGAAAACUACCAAGACAGAAGUUAGAUCUAAGCAUCUUAGAAAAGGCGAUAGUGGGAUAGUUCUCCU
                           750                           780                           810                           840
                     220                           230                           240                           250
 T  Q  K  I  K  G  H  L  S  G  V  R  D  E  F  F  I  V  R  G  T  C  D  D  S  L  L  E  A  R  A  R  F  S  Q  S  I  T  L  R
GACACAGAAGAUAAAAGGACAUCUCAGCGGAGUGCGGGAUGAAUUCUUUAUUGUUAGAGGAACGUGUGAUGAUAGUUUGUUGGAGGCAAGAGCUAGAUUCAGUCAAUCCAUCACACUGCG
                           870                           900                           930                           960
                     260                           270                           280                           290
 A  T  H  F  S  T  G  D  I  F  W  K  G  F  N  A  S  F  Q  E  Q  K  A  I  G  L  D  H  T  C  T  S  D  L  P  V  E  A  C  G
UGCGACUCACUUCUCAACUGGUGAUAUAUUUUGGAAGGGCUUUAACGCAUCCUUCCAGGAACAAAAAGCUAUAGGACUCGAUCACACUUGUACAUCUGACUUGCCAGUGGAAGCUUGUGG
                           990                          1020                          1050                          1080
                     300                           310                           320                           330
 H  V  A  A  L  M  C  Q  S  L  F  P  C  G  K  I  T  C  K  R  C  I  A  N  L  S  N  L  D  F  D  T  F  S  E  L  Q  G  D  R
GCAUGUUGCAGCUUUAAUGUGUCAAAGCUUAUUUCCGUGCGGAAAAUCACAUGUAAGAGAUGCAUAGCAAAUCUUAGUAACUUGGAUUUUGACACAUUUUCUGAAUUACAAGGUGAUAG
                          1110                          1140                          1170                          1200
                     340                           350                           360                           370
 A  M  R  I  L  D  V  M  R  A  R  F  P  S  F  T  H  T  I  R  F  L  H  D  L  F  T  Q  R  R  V  T  N  P  N  T  A  A  F  R
AGCAAUGCGAAUUCUGGAUGUAAUGAGAGCACGGUUUCCAAGUUUUACCCACACAAUUCGCUUCUUACACGACUUGUUCACUCAACGACGAGUCACUAAUCCCAACACUGCCGCUUUCAG
                          1230                          1260                          1290                          1320
                     380                           390                           400                           410
 E  I  L  R  L  I  G  D  R  N  E  A  P  F  A  H  V  N  R  L  N  E  I  L  L  L  G  S  K  A  N  P  D  S  L  A  K  A  S  D
GGAAAUUCUUCGUUUGAUUGGAGAUAGAAAUGAAGCACCAUUUGCGCAUGUCAAUCGAUUGAAUGAAAUUCUUUUACUUGGUUCAAAAGCCAAUCCUGACAGCCUUGCAAAGGCGUCAGA
                          1350                          1380                          1410                          1440
                     420                           430                           440                           450
 S  L  L  E  L  A  R  Y  L  N  N  R  T  E  N  I  R  N  G  S  L  K  H  F  R  N  K  I  S  S  K  A  H  S  N  L  A  L  S  C
CUCUUUGCUUGAGUUGGCGAGGUAUCUGAACAAUCGCACAGAGAACAUUCGCAACGGAUCAUUGAAGCAUUUCAGAAAUAAAAUUUCCUCAAAAGCGCAUUCAAAUCUGGCUUUGUCUUG
                          1470                          1500                          1530                          1560
                     460                           470                           480                           490
 D  N  Q  L  D  Q  N  G  N  F  L  W  G  L  A  G  I  A  A  K  R  F  L  N  N  Y  F  E  T  I  D  P  E  Q  G  Y  D  K  Y  V
CGACAAUCAGCUGGACCAGAAUGGAAAUUUUUUAUGGGGACUCGCCGGUAUUGCCGCAAAGAGAUUUCUCAAUAAUUACUUUGAAACAAUUGAUCCUGAGCAAGGUUAUGAUAAGUAUGU
                          1590                          1620                          1650                          1680
                     500                           510                           520                           530
 I  R  K  N  P  N  G  E  R  K  L  A  I  G  N  F  I  I  S  T  N  L  E  K  L  R  D  Q  L  E  G  E  S  I  A  R  V  G  I  T
CAUCCGGAAAAACCCAAAUGGAGAACGCAAAUUAGCAAUUGGCAAUUUUAUCAUUUCAACAAACCUUGAAAAGUUGCGCGAUCAACUAGAAGGGGAGUCGAUUGCACGGGUCGGAAUUAC
                          1710                          1740                          1770                          1800
                     540                           550                           560                           570
 E  E  C  V  S  R  K  D  G  N  Y  R  Y  P  C  C  C  V  T  L  E  D  G  S  P  M  Y  S  E  L  K  M  P  T  K  N  H  L  V  I
AGAGGAAUGUGUUAGUCGGAAGGAUGGUAAUUAUAGGUACCCAUGCUGCUGCGUCACUCUCGAAGAUGGUAGUCCAAUGUACUCAGAGCUUAAAAUGCCAACGAAAAAUCAUCUAGUAAU
                          1830                          1860                          1890                          1920
                     580                           590                           600                           610
 G  N  S  G  D  P  K  Y  L  D  L  P  G  E  I  S  N  L  M  Y  I  A  K  E  G  Y  C  Y  I  N  I  F  L  A  M  L  V  N  V  D
UGGCAAUUCAGGGGAUCCGAAAUACUUGGAUCUACCAGGUGAAAUUAGCAAUCUUAUGUACAUAGCAAAGGAAGGAUAUUGUUAUAUCAACAUAUUUCUUGCAAUGCUUGUUAAUGUUGA
                          1950                          1980                          2010                          2040
                     620                           630                           640                           650
 E  A  N  A  K  D  F  T  K  R  V  R  D  E  S  V  Q  K  L  G  K  W  P  S  L  I  D  V  A  T  E  C  A  L  L  S  T  Y  Y  P
UGAAGCUAACGCCAAGGACUUUACUAAGAGAGUGAGGGACGAGUCAGUACAAAAGCUUGGAAAGUGGCCAAGUUUAAUAGAUGUCGCAACUGAAUGUGCCUUACUAUCUACAUAUUAUCC
                          2070                          2100                          2130                          2160
                     660                           670                           680                           690
 A  A  A  S  A  E  L  P  R  L  L  V  D  H  A  Q  K  T  I  H  V  V  D  S  Y  G  S  L  N  T  G  Y  H  I  L  K  A  N  T  V
UGCGGCGGCUAGUGCAGAACUACCCAGGCUUCUAGUAGAUCAUGCUCAAAAGACAAUUCACGUUGUGGAUUCUUAUGGGUCGUUAAAUACGGGAUACCACAUCCUGAAAGCAAAUACAGU
                          2190                          2220                          2250                          2280
                     700                           710                           720                           730
 S  Q  L  E  K  F  A  S  N  T  L  E  S  P  M  A  Q  Y  K  V  G  G  L  V  Y  S  E  N  N  D  A  S  A  V  K  A  L  T  Q  A
GAGCCAACUUGAAAAGUUUGCUAGCAACACGCUAGAAUCACCAAUGGCACAAUAUAAAGUGGGUGGUCUCGUGUAUUCAGAGAACAAUGAUGCCAGUGCAGUCAAGGCAUUAACACAGGC
                          2310                          2340                          2370                          2400
                     740                           750                           760                           770
 I  F  R  P  D  V  L  S  E  L  I  E  K  E  P  Y  L  M  V  F  A  L  V  S  P  G  I  L  M  A  M  S  N  S  G  A  L  E  F  G
AAUAUUUCGGCCAGAUGUCUUAAGCGAAUUGAUAGAAAAAGAGCCUUAUCUUAUGGUCUUUGCCUUAGUAUCACCUGGGAUCUUAAUGGCAAUGUCAAAUAGUGGUGCACUCGAAUUUGG
                          2430                          2460                          2490                          2520
                     780                           790                           800                           810
 I  S  K  W  I  S  S  D  H  S  L  V  R  M  A  S  I  L  K  T  L  A  S  K  V  S  V  A  D  T  L  A  L  Q  K  H  I  M  R  Q
AAUUUCAAAAUGGAUUUCAAGCGAUCAUAGUCUGGUUCGAAUGGCAUCUAUUUUGAAAACUCUUGCCAGUAAAGUUAGUGUGGCUGAUACACUAGCUUUGCAAAAACACAUAAUGAGGCA
                          2550                          2580                          2610                          2640
                     820                           830                           840                           850
 N  A  N  F  L  C  G  E  L  I  N  G  F  Q  K  K  K  S  Y  T  H  A  T  R  F  L  L  M  I  S  E  E  N  E  M  D  D  P  V  L
GAAUGCAAAUUUUCUAUGUGGAGAGUUGAUAAAUGGCUUUCAGAAGAAGAAAUCGUAUACACAUGCAACGCGAUUUCUACUGAUGAUUUCAGAGGAAAAUGAAAUGGAUGAUCCUGUACU
                          2670                          2700                          2730                          2760
                     860                           870                           880                           890
 N  A  G  Y  R  V  L  E  A  S  S  H  E  I  M  E  K  T  Y  L  A  L  L  E  T  S  W  S  D  L  S  L  Y  G  K  F  K  S  I  W
GAAUGCAGGAUAUAGAGUAUUAGAAGCAUCGUCUCAUGAGAUAAUGGAAAAAACCUAUCUCGCACUGUUAGAGACAUCUUGGUCAGACUUAAGCUUGUAUGGAAAAUUCAAGUCAAUCUG
                          2790                          2820                          2850                          2880
                     900                           910                           920                           930
 F  T  R  K  H  F  G  R  Y  K  A  E  L  F  P  K  E  Q  T  D  L  Q  G  R  Y  S  N  S  L  R  F  H  Y  Q  S  T  L  K  R  L
GUUUACGCGAAAGCACUUUGGAAGAUACAAAGCAGAAUUGUUCCCAAAAGAGCAGACAGACUUGCAAGGACGCUACAGCAACUCGUUGCGGUUUCAUUACCAGAGUACGCUCAAGCGCUU
                          2910                          2940                          2970                          3000
                     940                           950                           960                           970
 R  N  K  G  S  L  C  R  E  R  F  L  E  S  I  S  S  A  R  R  R  T  T  C  A  V  F  S  L  L  H  K  A  F  P  D  V  L  K  F
GAGAAACAAGGGGAGUCUGUGUCGAGAAAGAUUUUUGGAAAGCAUUUCAAGUGCAAGACGCAGGACAACAUGUGCAGUUUUUAGUCUUUUGCACAAAGCAUUUCCGGAUGUUUUGAAAUU
                          3030                          3060                          3090                          3120
                     980                           990                          1000                          1010
 I  N  T  L  V  I  V  S  L  S  M  Q  I  Y  Y  M  L  V  A  I  I  H  E  H  R  A  A  K  I  K  S  A  Q  L  E  E  R  V  L  E
CAUUAACACAUUAGUCAUAGUUAGUUUGUCUAUGCAAAUAUAUUAUAUGCUUGUAGCAAUAAUACACGAGCAUAGGGCUGCAAAGAUCAAGAGCGCGCAGCUCGAAGAAAGAGUGUUAGA
                          3150                          3180                          3210                          3240
                    1020                          1030                          1040                          1050
 D  K  T  M  L  L  Y  D  D  F  K  A  K  L  P  E  G  S  F  E  E  F  L  E  Y  T  R  Q  R  D  K  E  V  Y  E  Y  L  M  M  E
GGAUAAGACCAUGCUAUUAUAUGAUGAUUUUAAAGCCAAGUUGCCAGAGGGGUCUUUUGAGGAGUUUUUGGAAUACACUCGACAGCGUGAUAAAGAAGUGUAUGAGUACCUGAUGAUGGA
                          3270                          3300                          3330                          3360
                    1060                          1070                          1080                          1090
 T  T  E  I  V  E  F  Q  A  K  N  T  G  Q  A  S  L  E  R  I  I  A  F  V  S  L  T  L  M  L  F  D  N  E  R  S  D  C  V  Y
AACAACUGAGAUUGUGGAAUUCCAAGCUAAAAACACGGGCCAAGCUAGCUUGGAAAGGAUUAUCGCAUUUGUGUCAUUAACACUAAUGUUAUUUGACAAUGAGCGUAGUGAUUGCGUGUA
                          3390                          3420                          3450                          3480
                    1100                          1110                          1120                          1130
 K  I  L  T  K  F  K  G  I  L  G  S  V  E  N  N  V  R  F  Q  S  L  D  T  I  V  P  T  Q  E  E  K  N  M  V  I  D  F  E  L
CAAAAUUCUGACAAAAUUUAAAGGCAUACUUGGUAGUGUGGAAAACAAUGUCAGAUUUCAGAGUCUAGAUACAAUAGUCCCAACUCAAGAGGAGAAGAACAUGGUAAUAGACUUUGAGCU
                          3510                          3540                          3570                          3600
                    1140                          1150                          1160                          1170
 D  S  D  T  A  H  T  P  Q  M  Q  E  Q  T  F  S  D  W  W  S  N  Q  I  A  N  N  R  V  V  P  H  Y  R  T  E  G  Y  F  M  Q
UGAUAGCGACACAGCACACACGCCGCAAAUGCAAGAGCAAACAUUCAGCGACUGGUGGAGCAACCAAAUAGCAAAUAAUCGCGUAGUUCCUCACUACAGAACAGAAGGCUAUUUCAUGCA
                          3630                          3660                          3690                          3720
                    1180                          1190                          1200                          1210
 F  T  R  N  T  A  S  A  V  S  H  Q  I  A  H  N  E  H  K  D  I  I  L  M  G  A  V  G  S  G  K  S  T  G  L  P  T  N  L  C
GUUCACCCGUAAUACGGCAUCAGCUGUUUCACAUCAAAUAGCCCACAAUGAACAUAAAGAUAUUAUUUUGAUGGGAGCAGUCGGUUCAGGAAAGUCAACGGGUCUGCCGACAAAUUUAUG
                          3750                          3780                          3810                          3840
                    1220                          1230                          1240                          1250
  K  F  G  G  V  L  L  L  E  P  T  R  P  L  A  E  N  V  T  K  Q  M  R  G  S  P  F  F  A  S  P  T  L  R  M  R  N  L  S  T
CAAAUUUGGCGGAGUGCUAUUACUUGAGCCCACGCGUCCACUCGCAGAGAAUGUGACAAAGCAAAUGAGGGGAAGCCCUUUCUUCGCAUCCCCAACUCUUAGAAUGCGCAAUCUCAGCAC
                          3870                          3900                          3930                          3960
                    1260                          1270                          1280                          1290
  F  G  S  S  P  I  T  V  M  T  T  G  F  A  L  H  F  F  A  N  N  V  K  E  F  D  R  Y  Q  F  I  I  F  D  E  F  H  V  L  D
AUUCGGCUCUAGCCCGAUUACUGUUAUGACAACUGGCUUUGCUUUACAUUUCUUUGCAAACAAUGUGAAGGAAUUUGAUCGGUACCAAUUUAUAAUAUUCGAUGAGUUCCAUGUGCUAGA
                          3990                          4020                          4050                          4080
                    1300                          1310                          1320                          1330
  S  N  A  I  A  F  R  N  L  C  H  E  Y  S  Y  N  G  K  I  I  K  V  S  A  T  P  P  G  R  E  C  D  L  T  T  Q  Y  P  V  E
UAGUAAUGCAAUAGCUUUCAGGAACCUUUGUCAUGAGUAUAGCUACAAUGGGAAGAUCAUAAAAGUCUCCGCAACUCCUCCAGGAAGAGAGUGUGAUUUAACCACUCAAUAUCCAGUUGA
                          4110                          4140                          4170                          4200
                    1340                          1350                          1360                          1370
  L  L  I  E  E  Q  L  S  L  R  D  F  V  D  A  Q  G  T  D  A  H  A  D  V  V  K  K  G  D  N  I  L  V  Y  V  A  S  Y  N  E
GUUACUAAUUGAGGAGCAGCUUAGCCUUCGGGACUUCGUUGAUGCACAAGGCACAGAUGCCCAUGCAGAUGUGGUUAAAAAGGGCGACAACAUUCUUGUGUACGUUGCAAGUUACAAUGA
                          4230                          4260                          4290                          4320
                    1380                          1390                          1400                          1410
  V  D  Q  L  S  K  M  L  N  E  R  G  F  L  V  T  K  V  D  G  R  T  M  K  L  G  G  V  E  I  I  T  K  G  S  S  I  K  K  H
AGUUGAUCAGUUGAGCAAAAUGUUGAAUGAGCGAGGUUUUCUGGUGACGAAAGUUGAUGGCAGAACUAUGAAGCUAGGAGGGGUUGAGAUCAUAACAAAAGGAAGCUCAAUCAAGAAACA
                          4350                          4380                          4410                          4440
                    1420                          1430                          1440                          1450
  F  I  V  A  T  N  I  I  E  N  G  V  T  L  D  V  D  V  V  V  D  F  G  L  K  V  V  P  N  L  D  S  D  N  R  L  V  S  Y  C
CUUCAUCGUGGCAACCAAUAUAAUCGAGAAUGGGGUAACACUUGAUGUCGAUGUUGUGGUAGACUUUGGACUCAAAGUCGUUCCAAAUCUUGAUUCCGAUAAUCGCUUAGUUAGUUAUUG
                          4470                          4500                          4530                          4560
                    1460                          1470                          1480                          1490
  K  I  P  I  S  L  G  E  R  I  Q  R  F  G  R  V  G  R  N  K  P  G  V  A  L  R  I  G  E  T  I  K  G  L  V  E  I  P  S  M
CAAAAUUCCCAUAAGCUUAGGUGAGAGAAUUCAAAGAUUUGGCCGAGUUGGUCGCAAUAAACCCGGAGUGGCGCUUAGAAUUGGGGAGACAAUAAAAGGGUUGGUGGAAAUUCCAUCUAU
                          4590                          4620                          4650                          4680
                    1500                          1510                          1520                          1530
  I  A  T  E  A  A  F  L  C  F  V  Y  G  L  P  V  T  T  Q  N  V  S  T  S  I  L  S  Q  V  S  V  R  Q  A  R  V  M  C  Q  F
GAUUGCAACAGAGGCGGCUUUUCUGUGUUUUGUUUACGGCUUACCAGUCACAACUCAGAAUGUCUCAACUAGCAUUUUAUCACAAGUGAGUGUUCGCCAAGCGCGAGUUAUGUGUCAAUU
                          4710                          4740                          4770                          4800
                    1540                          1550                          1560                          1570
  E  L  P  I  F  Y  T  A  H  L  V  R  Y  D  G  A  M  H  P  A  I  H  N  A  L  K  R  F  K  L  R  D  S  E  I  N  L  N  T  L
UGAACUCCCAAUCUUUUACACAGCUCAUUUGGUACGGUAUGAUGGGGCCAUGCAUCCAGCCAUUCACAAUGCACUAAAGCGAUUUAAGCUGCGAGAUAGCGAAAUCAACCUGAACACAUU
                          4830                          4860                          4890                          4920
                    1580                          1590                          1600                          1610
  A  I  P  T  S  S  S  K  T  W  Y  T  G  K  C  Y  K  Q  L  V  G  R  L  D  I  P  D  E  I  K  I  P  F  Y  T  K  E  V  P  E
GGCAAUCCCAACUAGCAGUUCAAAGACUUGGUAUACUGGAAAGUGUUACAAGCAAUUAGUGGGACGGCUGGACAUCCCUGAUGAGAUCAAGAUCCCAUUCUACACAAAGGAGGUACCUGA
                          4950                          4980                          5010                          5040
                    1620                          1630                          1640                          1650
  K  V  P  E  Q  I  W  D  V  M  V  K  F  S  S  D  A  G  F  G  R  M  T  S  A  A  A  C  K  V  A  Y  T  L  Q  T  D  I  H  S
AAAGGUGCCUGAGCAGAUCUGGGACGUUAUGGUGAAAUUUUCAAGUGACGCCGGCUUUGGGAGAAUGACAAGUGCCGCUGCAUGUAAGGUUGCAUACACACUGCAAACUGACAUUCAUUC
                          5070                          5100                          5130                          5160
                    1660                          1670                          1680                          1690
  I  Q  R  T  V  Q  I  I  D  R  L  L  E  N  E  M  K  K  R  N  H  F  N  L  V  V  N  Q  S  C  S  S  H  F  M  S  L  S  S  I
UAUACAGAGAACUGUGCAGAUUAUUGAUCGCCUACUUGAGAAUGAAAUGAAGAAGAGAAACCAUUUCAACUUGGUUGUAAAUCAAUCGUGUUCAUCUCACUUCAUGUCCUUAUCAUCAAU
                          5190                          5220                          5250                          5280
                    1700                          1710                          1720                          1730
  M  A  S  L  R  A  H  Y  A  K  N  H  T  G  Q  N  I  E  I  L  Q  K  A  K  A  Q  L  L  E  F  S  N  L  A  I  D  P  S  T  T
UAUGGCAUCACUCAGAGCCCAUUAUGCCAAGAAUCACACCGGGCAAAAUAUUGAAAUUCUUCAGAAGGCAAAGGCCCAGCUGCUUGAAUUCUCCAACCUUGCAAUAGACCCAUCAACAAC
                          5310                          5340                          5370                          5400
                    1740                          1750                          1760                          1770
  E  A  L  R  D  F  G  Y  L  E  A  V  R  F  Q  S  E  S  E  M  A  R  G  L  K  L  S  G  H  W  K  W  S  L  I  S  R  D  L  I
AGAAGCCUUGCGAGACUUCGGAUAUCUUGAGGCAGUAAGAUUCCAGAGUGAAUCGGAGAUGGCUCGUGGUCUUAAGUUAAGUGGACAUUGGAAGUGGUCACUCAUUAGUAGAGAUCUCAU
                          5430                          5460                          5490                          5520
                    1780                          1790                          1800                          1810
  V  V  S  G  V  G  I  G  L  G  C  M  L  W  Q  F  F  K  E  K  M  H  E  P  V  K  F  Q  G  K  S  R  R  R  L  Q  F  R  K  A
UGUUGUAAGUGGCGUCGGAAUUGGUCUUGGUUGCAUGUUGUGGCAAUUCUUCAAAGAGAAGAUGCAUGAACCAGUUAAAUUUCAAGGCAAGAGUAGACGCCGACUUCAAUUCAGAAAAGC
                          5550                          5580                          5610                          5640
                    1820                          1830                          1840                          1850
  R  D  D  K  M  G  Y  I  M  H  G  E  G  D  T  I  E  H  F  F  G  A  A  Y  T  K  K  G  K  S  K  G  K  T  H  G  A  G  T  K
AAGGGACGAUAAGAUGGGUUAUAUCAUGCAUGGUGAAGGGGACACAAUUGAACAUUUCUUUGGCGCGGCCUAUACAAAGAAAGGAAAGUCCAAAGGGAAAACUCAUGGCGCUGGAACGAA
                          5670                          5700                          5730                          5760
                    1860                          1870                          1880                          1890
  A  H  K  F  V  N  M  Y  G  V  S  P  D  E  Y  S  Y  V  R  Y  L  D  P  V  T  G  A  T  L  D  E  S  P  M  T  D  L  N  I  V
GGCGCACAAAUUUGUGAAUAUGUAUGGAGUCAGUCCUGAUGAAUAUUCAUAUGUACGUUAUCUAGAUCCAGUCACGGGUGCGACUCUAGAUGAAUCUCCAAUGACAGACUUAAACAUUGU
                          5790                          5820                          5850                          5880
                    1900                          1910                          1920                          1930
  Q  E  H  F  G  E  I  R  R  E  A  I  L  A  D  A  M  S  P  Q  Q  R  N  K  G  I  Q  A  Y  F  V  R  N  S  T  M  P  I  L  K
GCAAGAACACUUUGGAGAAAUCAGGAGGGAAGCUAUACUUGCUGAUGCAAUGUCACCACAACAAAGGAACAAGGGAAUCCAGGCAUACUUUGUUAGAAAUUCAACAAUGCCAAUUCUCAA
                          5910                          5940                          5970                          6000
                    1940                          1950                          1960                          1970
  V  D  L  T  P  H  I  P  L  K  V  C  E  S  N  N  I  A  G  F  P  E  R  E  G  E  L  R  R  T  G  P  T  E  T  L  P  F  D  A
AGUUGAUCUGACACCACACAUUCCUCUUAAAGUAUGUGAGUCUAAUAAUAUUGCUGGCUUUCCAGAGAGAGAAGGGGAAUUGCGGAGAACAGGCCCAACAGAAACACUCCCCUUUGAUGC
                          6030                          6060                          6090                          6120
                    1980                          1990                          2000                          2010
  L  P  P  E  K  Q  E  V  A  F  E  S  K  A  L  L  K  G  V  R  D  F  N  P  I  S  A  C  V  W  L  L  E  N  S  S  D  G  H  S
ACUGCCCCCAGAAAAACAAGAAGUUGCAUUCGAGUCAAAGGCGCUACUUAAGGGAGUGCGAGAUUUUAAUCCAAUCUCUGCUUGCGUAUGGCUCCUUGAGAACUCCUCGGAUGGGCAUAG
                          6150                          6180                          6210                          6240
                    2020                          2030                          2040                          2050
  E  R  L  F  G  I  G  F  G  P  Y  I  I  A  N  Q  H  L  F  R  R  N  N  G  E  L  T  I  K  T  M  H  G  E  F  K  V  K  N  S
UGAGAGACUGUUUGGCAUUGGUUUUGGCCCAUAUAUCAUUGCCAACCAACAUCUUUUUAGAAGGAACAAUGGAGAGUUGACUAUCAAAACCAUGCAUGGUGAGUUCAAAGUCAAGAACUC
                          6270                          6300                          6330                          6360
                    2060                          2070                          2080                          2090
  T  Q  L  Q  M  K  P  V  E  G  R  D  I  I  V  I  K  M  A  K  D  F  P  P  F  P  Q  K  L  K  F  R  Q  P  T  I  K  D  R  V
AACACAAUUGCAGAUGAAACCAGUUGAGGGCAGAGACAUAAUAGUUAUCAAAAUGGCUAAGGACUUCCCACCAUUCCCUCAAAAACUAAAAUUCAGACAGCCUACCAUCAAAGAUAGAGU
                          6390                          6420                          6450                          6480
                    2100                          2110                          2120                          2130
  C  M  V  S  T  N  F  Q  Q  K  S  V  S  S  L  V  S  E  S  S  H  I  V  H  K  E  D  T  S  F  W  Q  H  W  I  T  T  K  D  G
GUGCAUGGUAUCCACAAAUUUUCAGCAGAAAAGUGUCUCUAGUCUAGUGUCUGAGUCAUCACACAUUGUGCAUAAAGAGGACACUUCAUUCUGGCAACACUGGAUAACAACAAAGGAUGG
                          6510                          6540                          6570                          6600
                    2140                          2150                          2160                          2170
  Q  C  G  S  P  L  V  S  I  I  D  G  N  I  L  G  I  H  S  L  T  H  T  T  N  G  S  N  Y  F  V  E  F  P  E  K  F  V  A  T
ACAAUGUGGAAGUCCGCUGGUUUCAAUCAUUGAUGGAAAUAUUUUGGGGAUCCACAGCCUGACGCAUACGACCAAUGGUAGCAAUUACUUCGUGGAAUUUCCUGAGAAGUUCGUAGCUAC
                          6630                          6660                          6690                          6720
                    2180                          2190                          2200                          2210
  Y  L  D  A  A  D  G  W  C  K  N  W  K  F  N  A  D  K  I  S  W  G  S  F  T  L  V  E  D  A  P  E  D  D  F  M  A  K  K  T
AUAUCUUGAUGCCGCUGAUGGUUGGUGCAAGAAUUGGAAGUUCAAUGCUGAUAAGAUCAGUUGGGGUUCCUUUACAUUAGUUGAGGAUGCGCCCGAAGAUGACUUCAUGGCCAAGAAAAC
                          6750                          6780                          6810                          6840
                    2220                          2230                          2240                          2250
  V  A  A  I  M  D  D  L  V  R  T  Q  G  E  K  R  K  W  M  L  E  A  A  H  T  N  I  Q  P  V  A  H  L  Q  S  Q  L  V  T  K
UGUUGCCGCCAUCAUGGACGAUUUGGUCCGCACUCAAGGGGAGAAGCGAAAAUGGAUGUUGGAAGCAGCGCAUACAAAUAUUCAACCAGUUGCGCAUUUGCAAAGUCAGUUGGUCACAAA
                          6870                          6900                          6930                          6960
                    2260                          2270                          2280                          2290
  H  I  V  K  G  R  C  K  M  F  A  L  Y  L  Q  E  N  A  D  A  R  D  F  F  K  S  F  M  G  A  Y  G  P  S  H  L  N  K  E  A
GCACAUCGUGAAAGGCCGCUGUAAAAUGUUUGCACUUUAUCUUCAGGAAAAUGCAGAUGCGCGUGAUUUCUUUAAGUCAUUCAUGGGGGCAUAUGGUCCGAGCCACUUGAACAAGGAAGC
                          6990                          7020                          7050                          7080
                    2300                          2310                          2320                          2330
  Y  I  K  D  I  M  K  Y  S  K  Q  I  V  V  G  S  V  D  C  D  T  F  E  S  S  L  K  V  L  S  R  K  M  K  E  W  G  F  E  N
AUACAUCAAGGACAUCAUGAAGUAUUCCAAGCAGAUUGUGGUAGGCUCUGUUGAUUGUGACACAUUCGAAUCGUCAUUGAAAGUUCUUAGCAGGAAGAUGAAAGAGUGGGGAUUUGAGAA
                          7110                          7140                          7170                          7200
                    2340                          2350                          2360                          2370
  L  E  Y  V  T  D  E  Q  T  I  K  N  A  L  N  M  D  A  A  V  G  A  L  Y  S  G  K  K  K  Q  Y  F  E  D  L  S  D  D  A  V
UCUUGAAUAUGUCACAGAUGAGCAAACUAUCAAAAAUGCAUUAAACAUGGAUGCUGCAGUUGGUGCACUUUACAGCGGAAAGAAAAAGCAAUAUUUUGAGGACUUAUCUGAUGACGCUGU
                          7230                          7260                          7290                          7320
                    2380                          2390                          2400                          2410
  A  N  L  V  Q  K  S  C  L  R  L  F  K  N  K  L  G  V  W  N  G  S  L  K  A  E  L  R  P  F  E  K  L  I  E  N  K  T  R  T
AGCAAAUCUGGUUCAAAAGAGUUGUCUCCGCUUAUUCAAAAAUAAACUAGGCGUCUGGAAUGGAUCACUUAAAGCGGAACUUAGACCGUUCGAGAAGCUCAUCGAGAAUAAAACACGGAC
                          7350                          7380                          7410                          7440
                    2420                          2430                          2440                          2450
  F  T  A  A  P  I  E  T  L  L  G  G  K  V  C  V  D  D  F  N  N  H  F  Y  S  K  H  I  Q  C  P  W  S  V  G  M  T  K  F  Y
AUUCACAGCAGCACCAAUUGAGACAUUGCUCGGAGGAAAAGUCUGUGUUGAUGAUUUUAACAAUCAUUUCUACAGCAAGCACAUACAAUGCCCUUGGAGCGUUGGAAUGACAAAGUUCUA
                          7470                          7500                          7530                          7560
```

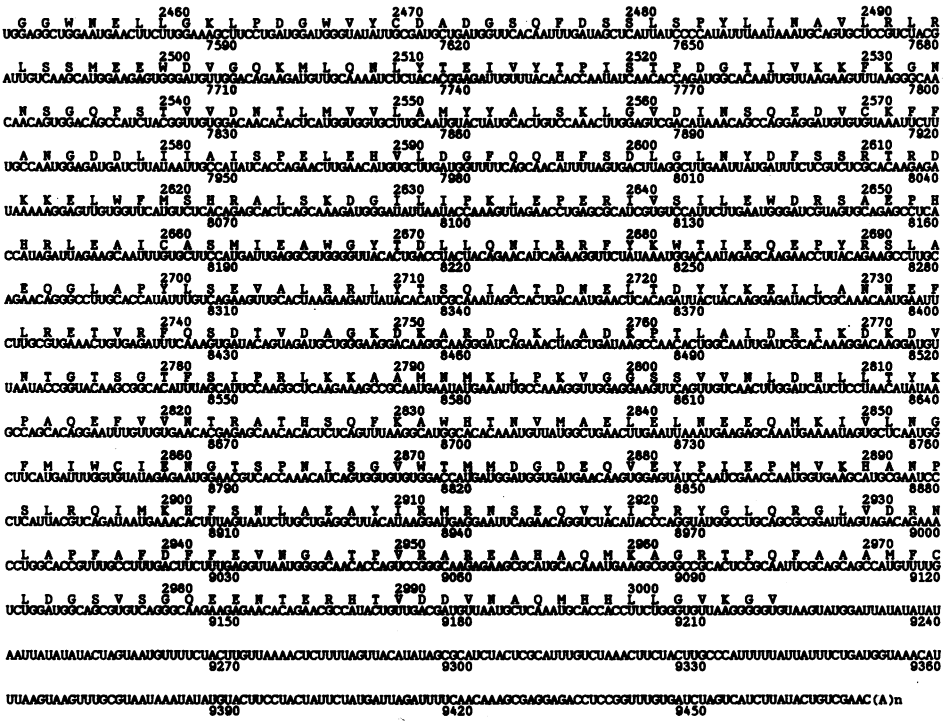

Fig 2. Nucleotide sequence of TVMV RNA and derived amino acid sequence of the putative polyprotein. The amino acid sequence of the open reading frame, which starts at nucleotide 206 is shown using the single letter code: A = Ala, C = Cys, D = Asp, E = Glu, F = Phe, G = Gly, H = His, I = Ile, K = Lys, L = Leu, M = Met, N = Asn, P = Pro, Q = Gln, R = Arg, S = Ser, T = Thr, V = Val, W = Trp, Y = Tyr. Numbers printed above the sequence identify amino acid residue positions. Numbers below the sequence identify nucleotide residue positions. The last digit of each number is aligned with the amino acid or nucleotide residue in question.

rhinovirus RNAs (34). Computer analysis showed that 63.5% of codons have either A or U at the third position.

Computer translation of the RNA sequence and its complement in all three reading frames revealed the presence of a single open reading frame of 9015 bases, starting at nucleotide position 206 (Fig. 3). A termination codon (UAA) occurs at position 9221. This open reading frame can encode a polypeptide of 3005 amino acid residues (340 kDa). These data demonstrate that our previous report of internal termination codons in the TVMV RNA sequence was incorrect (35). The region upstream from the predicted start of translation contains three presumably unused initiation codons. They are

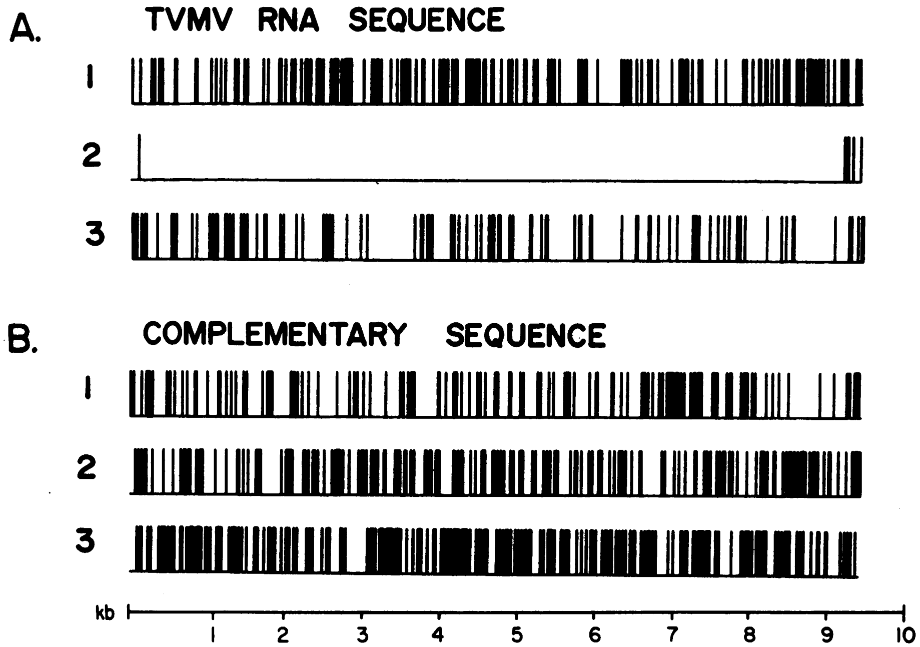

Fig 3. Termination Codons in TVMV RNA. Computer translation of the sequence in Fig. 2 was performed in all three reading frames for both the viral and complementary nucleotide sequence. The numbers on the left indicate the three different reading frames. Each vertical line represents the location of 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[a]	Chemically Determined[b]	Amino acid	Predicted[a]	Chemically Determined[b]
		residues per mole			
Ala	41	38	Ala	24	23
Arg	27	27	Arg	14	12
Asx	63	63	Asx	32	33
Cys	13	ND[c]	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

[a] Predicted from the nucleotide sequence.

[b] Calculated by amino acid analysis of the isolated proteins.

[c] Not determined

RNA, however, is found 94 bases form 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 procaryotes and eucaryotes 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).

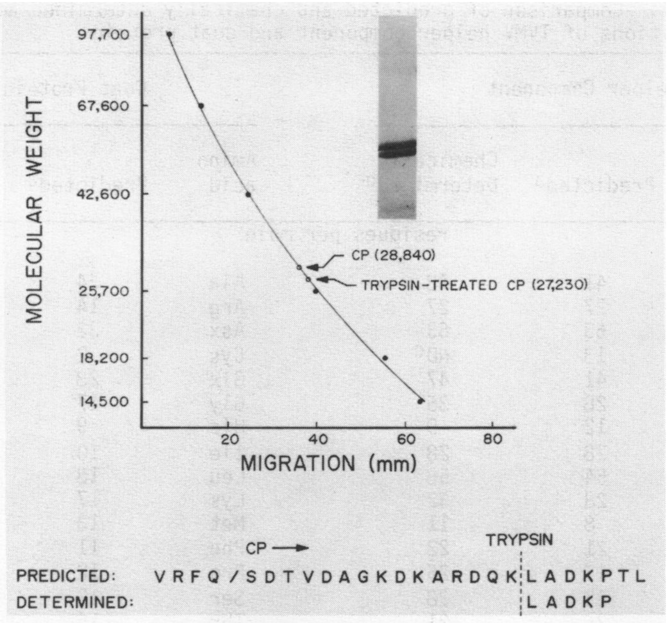

Fig 4. Determination of amino acid sequence of trypsin-treated TVMV coat protein (CP). Purified TVMV was treated with trypsin and analyzed by SDS-PAGE, with protein visualized by Coomassie blue staining (inset). The molecular weight of the resulting CP fragment was estimated by electrophoretic mobility. The amino acid sequence of the polyprotein in the region of the proposed cleavage site (/) separating CP from NI_b, 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 glutaminyl 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_a cistrons,

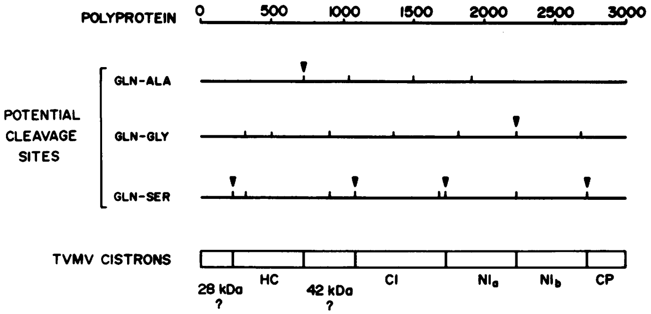

Fig 5. Proposed protease cleavage map of the TVMV polyprotein. Computer analysis was performed to localize the positions (vertical lines) of the dipeptides Gln-Ala, Gln-Gly, and Gln-Ser in the putative TVMV polyprotein sequence (horizontal lines). Those which represent the proposed cleavage sites for 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_a 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^{-3} Measured[a]	Literature References	MW X 10^{-3} Predicted[b]	Amino Acid Positions
HC	53	48	54.0	248-731
CI	70	16	71.1	1113-1747
NI_a	52	10	53.3	1748-2224
NI_b	56	10	59.1	2225-2740
CP	29	this report	29.6	2741-3005

[a] Estimated from SDS polyacrylamide gel electrophoresis in the indicated studies.

[b] 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.

[1] Supported by a grant from the Brown and Williamson Tobacco Corporation, Louisville, KY.

[2] Current address: Research Institute of Scripps Clinic, 10660 North Torry Pines Road, La Jolla, CA 92037.

[3] Current address: Department of Virus Research, John Innes Institute, Colney Lane, Norwich NR4 7UH, United Kingdom.

[4] 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. Siaw, 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.

22. Norrander, J. Kempe, T., Messing, J. (1983) Gene 26, 101-106.
23. Messing, J (1983) in Wu, R., Grossman, L., Moldave, K. (eds), Meth. Enzymol., Academic Press, Orlando, Fla., Vol. 101, pp. 20-89.
24. Sanger F., Nicklen, S., Coulson, A.R (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
25. Biggin, M.F., Gibson, T.J., Horig, G.G. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
26. Dale, R.M.K., McClure, B.A., Houchins, J.P. (1985) Plasmid 13, 31-40.
27. Gould, A.R., Symons, R.H. (1982) Eur. J . Biochem. 126, 217-226.
28. Queen, C., Korn, L.J. (1984) Nucleic Acids Res. 12, 581-599.
29. Thornbury, D.W., Hellmann, G.M., Rhoads, R.E., Pirone, T.P. (1985) Virology 144, 260-267.
30. Hirs, C.H.W., Stein, W.H., Moore, S. (1954) J. Biol. Chem. 211, 941-950.
31. Hewick, R.M., Hunkapiller, M.W., Hood, L.E., Dreyer, Q.J. (1981) J. Biol Chem. 256, 7990-7997.
32. van Wezenbeek, P., Verver, J., Harmsen, J., Vos, P., van Kammen, A. (1983) EMBO J. 26, 941-946.
33. Lomonossoff, G., Shanks, M. (1983) EMBO J. 2, 2253-2258.
34. Callahan, P.L., Mizutani, S., Colonno, R.J. (1985) Proc. Natl. Acad. Sci. USA 82, 732-736.
35. Hellmann, G.M., Shaw, J.G., Rhoads, R.E. (1985) Virology 143, 23-34.
36. Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.
37. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., Gralla, J. (1973) Nature (London) New Biol. 246, 40-41.
38. Yanofsky, C. (1981) Nature (London) 289, 751-758.
39. Birchmeir, C., Grosschedl, R., Birnstiel, M.L. (1982) Cell 28, 739-745.
40. Holmes, W.M., Platt, T., Rosenberg, M. (1983) Cell 32, 1029-1032.
41. Ahlquist, P., Dasgupta, R., Kaesberg, P. (1981) Cell 23, 183-189.
42. Larsen, G.R., Semler, B.L., Wimmer, E. (1981) J. Gen. Vir. 37, 328-335.
43. Howarth, A.J., Canton, J., Bossert, M., Goodman, R.M. (1985) Proc. Natl. Acad. Sci. USA 82, 3572-3576.
44. Hobom, G., Grosschedl, R., Lusky, M., Scherer, G., Schwarz, E., Kossel, H. (1978) Cold Spring Harbor Symp. Quant. Biol., Cold Spring Harbor, New York, Vol. 43, pp. 165-178.
45. Racaniello, V.R., Baltimore, D. (1981) Proc. Natl. Acad. Sci. USA 78, 4887-4891.
46. Zabel, P., Moerman, M., Lomonossoff, G., Shanks, M., Beyreuther, K. (1984) EMBO J. 3, 1629-1634.
47. Forss, S., Schaller, H. (1982) Nucleic Acids Res. 10, 6441-6450.
48. Thornbury, D.W., Pirone, T.P. (1983) Virology 125, 487-490.