Multiple proteins and subgenomic mRNAs may be derived from a single open reading frame on tobacco mosaic virus RNA

Tony Hunter, Richard Jackson* and David Zimmern +

The Salk Institute, P.O. Box 85800, San Diego, CA 92138, USA, *Biochemistry Department, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, and + MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received 27 October 1982; Revised and Accepted 3 January 1983

ABSTRACT

It has previously been shown that messenger activity for a protein of Mr ca. 30k exists in RNA fractions extracted from particles of either native or alkali stripped U1 TMV, or from cowpea strain TMV, that are smaller than full genomic length. Analysis of sucrose gradient fractions containing this activity reveals a number of slightly smaller template activities directing synthesis of proteins between 18.5k and 29k in size. All of these messenger activities, including that for the 30k protein, respond to cap analogues in anomalous ways. Discrete RNA species that include active mRNAs for these proteins can be demonstrated in the same fractions by labelling with preparations of vaccinia capping enzyme and [α-32P] GTP without prior β-elimination. Detailed analysis of three of these proteins (of Mr s ca. 30k, 29k and 23k) by peptide mapping and translation of purified vaccinia-labelled RNA demonstrates that all three are unrelated to the large early TMV proteins, but are related to each other in such a way as to form a nested set with staggered N termini and identical C termini. mRNAs of chain lengths ca. 1900 and 1500 bases direct synthesis of the 30k and 23k proteins respectively, an mRNA of about 1850 bases directs both 29k and (perhaps because of cross-contamination) 30k synthesis. Initiation codons for the 29k and 23k proteins have been mapped at positions 4960-4962 and 5191-5193 respectively on TMV RNA. Since all three encapsidated templates have similar properties we conclude that either there is a family of 30k-related proteins with unusual mRNAs, or that none of these in vitro translation products are directed by physiological templates.

INTRODUCTION

The tobacco mosaic virus genome comprises a single RNA molecule of 6.4 kb which encodes at least four proteins (reviewed in Ref. 1). Two large proteins which can be detected early in infection or translated from full length viral RNA in vitro are encoded in the 5' terminal 4.9 kb of the genome, the larger protein being produced by readthrough of the amber terminator at the C terminus of the shorter one. The viral coat protein is encoded within the 3' terminal 700 bases and its ribosome binding site is not accessible on intact RNA; a small functional mRNA species (a subgenomic RNA) is produced from this region of the genome late during infection. Since this RNA does not include
the virus assembly origin (located between 850 and 950 bases from the 3' end of the genome in the common strain of TMV) it is not packaged and little or no coat protein is detectable among the translation products of total RNA extracted from virions. Beachy and Zaitlin (2) reasoned that if other subgenomic RNAs were produced by the virus which did include the assembly origin, these mRNAs would be packaged and their translation products would appear among the products of RNA extracted from virus rods of less than the full genomic length. They demonstrated that synthesis of a protein of \( M_r = 30,000 \) was directed by short rod RNA from 3 strains of TMV, and showed that the template RNAs were probably 3' coterminal with both the genomic RNA and the coat protein mRNA. This protein has proved difficult to detect in vivo, although it probably corresponds to the 31k protein occasionally seen in small amounts in infected tobacco protoplasts by Beier et al. (3). The question of whether these four proteins exhausted the coding capacity of the TMV genome remained open until recently, since molecular weight estimates for the large early proteins could only be approximate. The entire sequence of TMV RNA has now been determined (4), showing that the three open reading frames expected from earlier work (one interrupted by an in phase amber codon) span the genome almost completely, and that the strand complementary to the known mRNAs has little coding capacity.

The work described here emerged from attempts to analyse the proteins translatable from short packaged viral RNA and their templates in more detail. We were initially concerned to establish that the 30k gene product was indeed distinct from the two early proteins, and to localise the gene and define its mRNA more precisely. In the course of these experiments, we discovered that the 30k protein was only one of a family of related overlapping proteins whose synthesis was directed by short packaged RNA. In this paper we show that there are at least three such proteins, each encoded by a separate RNA, and that they are distinct from the large early proteins but related to each other in such a way that they form a nested set with staggered N termini and identical C termini.

**MATERIALS AND METHODS**

**Viruses** Common tobacco strains of TMV were obtained from Milt Zaitlin (strain Ul) and Jo Butler (original source: UC Berkeley). Limited RNA sequence comparisons suggest these strains are identical. Cowpea strain TMV (C\(_c\) TMV free of the contaminating C\(_t\) strain) was also from Milt Zaitlin.

**In vitro translation** Reactions in either the wheat germ system (supplemented
with spermidine) or the mRNA dependent reticulocyte lysate (supplemented with 50 µg/ml Boehringer calf liver tRNA) were carried out as previously described (5,6). Final concentrations of reaction components were as follows: mRNA (in both systems) at 100 µg/ml, [35S] methionine (Amersham, specific activity >1000 Ci/m mole) at 200-1000 μCi/ml, the cap analogue m GTP at 80 μM in the wheat germ and at either 300 μM or 1 mM (together with an equal molarity of supplementary MgCl₂) in the mRNA dependent reticulocyte lysate, and the S-adenosyl methionine antagonist S-adenosyl homocysteine at 0.5 mM in the wheat germ. For both systems 10 µl reactions were incubated for 60 min. at 30°C, then treated with pancreatic RNAse and solubilised in gel sample buffer. The equivalent of 2 µl of each sample was analysed by electrophoresis on 10% or 15% SDS-polyacrylamide gels (7). Analytical gels were impregnated with diphenyloxazole and exposed to preflashed X-ray film (8). Preparative gels were dried directly and autoradiographed.

Alkaline stripping and preparation of protected RNA Partially stripped virion preparations (PSV) were made by dialysis against 10 mM sodium borate pH 9.2 for 15 hr. at 0°C, followed by digestion with micrococcal nuclease to remove the exposed RNA tails, inactivation of the nuclease with EGTA, and phenol extraction of protected RNA as previously described (9). A preparation of PSV RNA was sedimented on a 15-40% linear sucrose gradient containing 20 mM sodium acetate pH 5.2, 1 mM EDTA, 0.1% SDS for 13 hr. at 34,000 rpm at 23°C in an SW 41 rotor. Fractions sedimenting between approximately 20S and 8S were pooled, precipitated with ethanol, and used for translation experiments and as a starting material for further purification.

Gel electrophoresis of PSV RNA 20 µg of PSV RNA was denatured by boiling for 40 sec. and then electrophoresed in a single lane of a 2% acrylamide/0.5% agarose composite gel run in the E buffer of Loening (10,11). The track was sliced into 2.5 mm sections and the RNA recovered from each slice by chromatography on cellulose (11). The eluted RNA was precipitated with ethanol following the addition of 10 µg of carrier tRNA. One tenth of each RNA fraction was translated in the mRNA dependent reticulocyte lysate.

Vaccinia capping enzyme reactions Gradient fractions were assayed using a scaled down version (total volume 2-5 µl) of the published protocol (12,13). Final concentrations of reaction components were as follows: [α-32P] GTP (Amersham) at either 30 μM (specific activity 400 Ci/m mole) or 10 μM (specific activity 3000 Ci/m mole), S-adenosyl methionine, where present, 50 μM, ATP (to suppress terminal transferase activity (14)) 0.16 mM, RNA 0.1 to 3 mg/ml, enzyme (gift of B. Moss: 0.5 µl transferred 0.77 pmole GTP to β-eliminated TMV
RNA in 20 min) at 0.25 μl per 5 μl reaction.

**Purification of RNA substrates in above reaction** TMV virion RNA was prepared by phenol-chloroform-SDS extraction (15). 80-100 μg batches of RNA were heated to 60°C for 3 min. in 1% SDS, then fractionated on 5 to 25% linear sucrose gradients containing 0.1 M NaCl, 10 mM TrisCl, pH 7.4, 1 mM EDTA and 0.1% SDS for 2½ hrs at 15° and 48,000 rpm in an SW 50.1 rotor. Larger batches (ca. 1 mg) were treated similarly except that gradients were run in an SW 27 rotor at 22,000 rpm for 16-17 hrs. Fractions to be assayed were precipitated three times with ethanol, redissolved in water, and stored frozen. Aliquots (one-fifth to one-half fractions in a volume of 1 to 3 μl) were labelled as described above, an equal volume of formamide dye solution was added, and the reaction mixture heated to 80°C for 2 min., chilled, and layered on a 40 cm long 3% acrylamide thin gel (a standard sequencing gel except for halved concentrations of acrylamide and bisacrylamide). Samples were run into the gel at 40 mA and then run at 35 W constant power for 4 to 6 hr. (both dyes run off). Gels were autoradiographed at -70°C using preflashed Fuji X-ray film and Dupont Cronex intensifying screens (8). Bands were excised and the whole gel slices were soaked in 0.5 M sodium acetate, 1 mM EDTA, 0.1% SDS (0.25 ml/cm width of gel) for 6 hr. at room temperature. The eluted RNA was recovered by ethanol precipitation. In some experiments 15 μg yeast tRNA was added to each 0.25 ml elution, but this can be omitted. One-tenth to one-half aliquots were assayed for messenger activity in the mRNA-dependent reticulocyte lysate. Note: The above protocol has been used for scanning gradients. For preparative purposes much cleaner gels can be obtained by labelling sized RNA in bulk and resedimenting to remove unincorporated GTP.

**Peptide mapping** Proteins were recovered from preparative gels, oxidised with performic acid and subjected to tryptic digestion as described (17). Tryptic digests were resolved in two dimensions on cellulose thin layer plates by electrophoresis at pH 4.7 followed by ascending chromatography (18). Where indicated plates were dipped in molten 2-methylnaphthalene containing 0.4% diphenyloxazole to enhance the detection of radioactivity. Secondary digestion with *Staphylococcus aureus* V8 protease was carried out on tryptic digests after inactivation of the residual trypsin by boiling (19).

**Peptide analysis** Areas of cellulose corresponding to particular peptides were scraped from the thin layer plates and the radioactive peptides were eluted. Peptides were digested with 1 μl carboxypeptidase B (Sigma, 2 mg/ml, 135 units/mg) in 10 μl 50 mM ammonium bicarbonate, pH 7.8 for 25 min. at 30°C. The digests were analysed following lyophilisation. [14C] N-acetyl-Met-Glu
was synthesised by acetylating the oxidised form of the dipeptide Met-Glu by treatment with a 10-fold molar excess of $^{14}$C acetic anhydride (18.5 mCi/m mole) in 50% pyridine for 1 hr at 23°C. The acetylated peptide was purified by thin-layer electrophoresis at pH 4.7.

RESULTS

The major translation product of unfractionated RNA extracted from TMV particles is the large protein encoded in the left half of the genome, variously referred to as 110k, 130k or 140k by different investigators. (Since a precise molecular weight of 126k can now be deduced from the nucleotide sequence $^{4}$, we will refer to it in the remainder of the paper as pl26, and to its readthrough product as pl83, while retaining the convention of denoting proteins whose molecular weights are known only approximately from SDS gel measurements as the "-k" protein). Smaller products are also evident, among them the 30k band described by Beachy and Zaitlin $^{2}$. Our initial approach to characterising this protein and its template took advantage of the fact that in dilute, mildly alkaline solutions, TMV coat protein is stripped away from its RNA in a polar fashion$^{9}$, starting at the 5' end. This reaction can be used to generate 3' coterminus RNA fragments with a range of discrete sizes, denoted partially stripped virus RNAs (PSV-RNAs). Six PSV-RNAs are major products ranging from PSV1 RNA (full length or nearly full length RNA) to PSV6 RNA (approximately the 3' terminal one-sixth of the genome, from the viral assembly origin to the 3' end). The advantage of using PSV RNAs that are smaller than full length is that the major ribosome binding site 70 nucleotides from the 5' terminus has been deleted, allowing one to concentrate on independently initiated proteins. Preliminary experiments showed that synthesis of a 30k protein was indeed directed by sucrose gradient fractions containing the smaller PSV RNAs (q.v. ref. 21).

The 30k protein is distinct from the 110k and 165k proteins. Tryptic peptide maps were prepared of the 30k PSV-RNA product labelled with $^{35}$S methionine, and compared with maps of pl26 and pl83 (Fig. 1). Inspeclion showed that the 30k protein maps had few, if any, spots in common with the larger proteins, which were found to be highly related as expected, although the near-identity seen in Fig. 1 was surprising. Maps of mixture of tryptic digests of the 30k in vitro product and either pl26 or pl83 confirmed their lack of similarity. We concluded that the 30k protein was indeed the product of a distinct gene, confirming the results of Beachy and Zaitlin $^{2}$. We also repeated this experiment using another strain of TMV, the cowpea
Fig. 1: The tryptic peptide map of U1-TMV 30k protein is unrelated to those of pl26 or pl83.
The origins are in the bottom left hand corner. 1st dimension electrophoresis was from left to right, followed by ascending chromatography. c,d and e are spots found on the map of pl83, but not that of pl26. 30k peptides are arrowed in the mixture with a pl83 digest.
strain, in which the viral assembly origin is located about 500 bases closer to the 3′ end of the RNA than in the common (tobacco) strain and so is within the coat protein gene (22). In this case we used intact virions, not the products of alkaline stripping, to prepare RNA since the polarity of alkaline stripping of cowpea strain TMV has not been determined, and with another agent (SDS) stripping is known to proceed in the opposite orientation to that in common TMV (23). Again, the 30k and 130k proteins had dissimilar peptide maps, which were also dissimilar to those from the common strain (data not shown).

Smaller proteins related to the 30k protein from in vitro translation.

In addition to the 30k translation products, we noticed that synthesis of other polypeptides in the 18,000 to 30,000 molecular weight range was directed by the same sucrose gradient fractions containing short PSV or cowpea TMV RNAs (Fig. 2). Other authors (2,3,24) have also noted in vivo synthesis of

![Figure 2](image-url)

**Fig. 2:** (a) Comparison of translation products of size selected U1-PSV and cowpea strain RNAs in the wheat germ (WG) and messenger-dependent reticulocyte lysate (MDL) systems, and (b) Effects of m7GTP and S-adenosyl homocysteine (SAH) on translation of the RNAs of fig. 2(a) in the wheat germ system.

The samples in panels (a) and (b) were each run in parallel - intervening tracks containing translation products of larger and smaller sucrose gradient fractions have been removed for clarity.
several of these proteins in a variety of contexts. Most of these polypeptides were synthesised in both the wheat germ and messenger-dependent reticulocyte lysate (MDL) systems, although not always with equal efficiency, the main exception being a cowpea strain product of Mr=29k, which was not made in the MDL although it was a major product in the wheat germ. (The MDL also incorporates no $^{35}$S label into cowpea strain coat protein, but this seems simply to reflect inefficient removal of the N-terminal Met-Met sequence in the wheat germ, there being no internal methionine (T.H., unpublished results)). Although such proteins might be dismissed as artifacts resulting from premature termination or from translation of RNA fragments, some of them appeared in yields comparable to that of the Mr=30k protein and/or had similar molecular weights in both the PSV RNA and cowpea strain RNA translations, in particular two proteins of apparent Mr = 29,000 and 23,000. We therefore examined these proteins further to see if they might be genuine, conserved translation products with independent mRNAs.

Since it ought to be possible to distinguish translation directed by RNA fragments from translation of capped subgenomic RNAs on the basis of the ability of cap analogues such as $^7$GTP to inhibit translation of capped mRNAs, we first examined the effect of $^7$GTP on translation of the smaller proteins. When assayed in the wheat germ system $^7$GTP somewhat inhibited translation of both the 30k protein and the smaller proteins, but not to the same degree as cowpea strain coat protein (Fig. 2b). However, S-adenosyl homocysteine (SAH), an inhibitor of the endogenous wheat germ capping enzyme which was included in control reactions, unexpectedly enhanced translation of some of the proteins (eg the cowpea strain 30k and 29k proteins, Fig. 2b). Independently, Pelham, who had been investigating the possible activation of spurious ribosome binding sites on a number of RNAs by nuclease cleavage, reported that proteins of Mr = 30k, 29k and 23k were prominent among the products of nuclease treated TMV RNA (in contrast to coat protein, whose synthesis is not activated by nuclease treatment - T. Hunter, unpublished results), and were $^7$GTP resistant in the messenger dependent reticulocyte lysate. We subsequently confirmed this result for the proteins we had identified among the products of PSV-RNA translation. We found this result puzzling since none of the proteins, including the 30k protein, responded as if their messengers were capped, and yet the evidence suggested that the 30k protein was the product of a distinct gene and subgenomic mRNA (Ref. 2 and Fig. 1). Although the cap analogue results seemed contradictory and therefore inconclusive, the $^7$GTP resistance of these proteins when
translation is assayed in the MDL has proved a useful distinguishing feature, and we have used this system for all subsequent experiments.

As a second line of approach, we examined the sizes of the templates for translation of the 30k and associated proteins. If these proteins simply resulted from premature termination of 30k translation, then the same mRNA template should direct synthesis of all three proteins. However, when a sample of PSV-RNA was fractionated further by gel electrophoresis, and RNA eluted from sliced fractions was used to direct translation in the MDL system there were clearly separable messenger activities (Fig. 3). The band of Mr = 29,000 appeared in all the fractions showing 30k protein synthesis but messenger activity was somewhat enriched on the low MW side of the 30k mRNA.

Fig. 3: Fractionation of mRNA activities in U -PSV RNA.
A sample of PSV RNA giving total translation products in the two right-hand tracks was electrophoresed on an acrylamide agarose composite gel. Individual gel slices tested for template activity in the messenger dependent reticulocyte lysate are displayed with fractions from the top of the first gel on the left. Track C = no RNA. Positions of stained PSV RNAs are shown below the gel.
peak. At slightly lower MW, there was messenger activity for an Mr = 28,000 protein. The peak of messenger activity for the 23k protein was clearly distinguishable, but a fifth band of Mr = 20,000 also appeared in the same fractions, enriched on the lower MW side, while a sixth band of Mr = 18,500 appeared to be synthesised by RNA of still lower MW. This result was inconsistent with the idea that at least three of the smaller proteins were premature termination products of the larger ones on the same mRNA template. Furthermore, although the RNA separation gel was not denaturing, and perhaps because of this the template activities for each protein were distributed through several fractions, in no case did the peaks of template activity seem to coincide with major bands of PSV RNA detectable by staining (Fig. 3), suggesting that their templates were distinct minor RNAs which had cosedimented with the PSV RNAs. Indeed, in the converse experiment, when PSV RNA bands located by staining were directly assayed by elution and translation, the 30k polypeptide was synthesised preferentially by PSV 3 RNA, as expected from the experiment in Fig. 3, whereas on gradients the 30k messenger activity has been reported to cosediment with PSV RNAs 5 and 6 (which are in fact too short to include the 30k open reading frame). These observations suggested that the active mRNAs were minor species cofractionating with different PSV RNAs according to the method of purification.

Isolation and preliminary characterisation of the presumptive mRNA species.

In order to obtain more direct evidence for the existence of discrete mRNAs we attempted to isolate the putative messengers directly, using RNA preparations that had not been subjected to the alkaline stripping procedure to simplify interpretation. Our strategy for mRNA isolation was also initially based on the assumption that we were dealing with capped subgenomic mRNAs. Vaccinia virions contain enzymes which will cap and methylate di- or triphosphate 5' termini on RNAs. Vaccinia virions contain enzymes which will cap and methylate di- or triphosphate 5' termini on RNAs. However, we noticed that incorporation into control samples of TMV RNA that had not been β-eliminated, was significantly higher than the background expected, and further investigation showed that label was being incorporated into discrete RNA species. Moreover, when samples of native viral RNA (common tobacco strain of TMV) were fractionated on sucrose gradients before labelling, a distinct peak of substrate RNA could be found sedimenting at about 60% of the velocity of full length viral RNA. Analysis of these labelled species on a
3% acrylamide–8M urea thin gel is shown in Fig. 4. Although backgrounds in the gels were high, mainly due to the large quantity of unincorporated GTP run through the gel, a minimum of three distinct bands were visible, plus a number of less distinct candidates emerging from the background. Approximate size estimates of the three major RNAs (denoted bands 1, 2 and 3 in decreasing order of size) were close to those found for the messenger activities responsible for 30k, 29k and 23k synthesis in the experiment shown in Fig. 3 (Table 1). Since the RNA had been sedimented before labelling it seemed unlikely that degradation by the labelling procedure could account for the result, and it occurred to us that if the 30k mRNA, or a substantial proportion of it, were not capped, but di- or triphosphorylated instead, this might account for some of the contradictory results with cap analogues (see Discussion).

![Electrophoresis diagram](image)

**Fig. 4**: Separation and in vitro translation of [α-32P] GTP labelled RNAs. Top: 3% acrylamide–8M urea gel separation. Below: Fluorogram of translation products directed by the eluted RNA in the messenger-dependent reticulocyte lysate ± mGTP. Molecular weight markers on the left include glutamate dehydrogenase (56k), creatine kinase (40k) and soybean trypsin inhibitor (21k).
Table 1: Sizes of PSV mRNA activities (a) and of vaccinia labelled RNAs (b).

<table>
<thead>
<tr>
<th>Expt. no.:</th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markers:</td>
<td>mRNA,</td>
<td>mRNA,</td>
<td>CCMV</td>
<td>PSV RNAs 5&amp;6</td>
</tr>
<tr>
<td>PSV RNAs</td>
<td>2200</td>
<td>1950</td>
<td>1800</td>
<td>1850*</td>
</tr>
<tr>
<td>23k mRNA</td>
<td>1500</td>
<td>1430</td>
<td>1500</td>
<td>1570</td>
</tr>
</tbody>
</table>

*Corrected for 10% faster migration of rRNA due to higher G+C composition

Translational activity of the RNA bands eluted from the 3% urea gel was assayed in the MDL, and the results are also shown in Fig. 4. Band 1 directed the synthesis of a 30k protein and band 3 synthesis of a 23k protein, almost exclusively. Band 2 RNA, which is only slightly smaller than that directing 30k synthesis gave both a 30k and a 29k protein. Each RNA in the experiment shown also directed synthesis of a single protein in the 40k to 56k size range, but these have not been seen reproducibly.

These RNAs were capped using [α-32P]GTP, but not methylated since S-adenosyl methionine was omitted from the reaction mixture. Methylation of the cap is required for a major enhancement of the efficiency of translation in naturally capped RNAs (27,31) and so we expected that the intrinsic sensitivity of template activity of these RNAs to competition by cap analogues would be little affected by the partial modification, which experience with other substrates shows is only likely to be a maximum of 30% complete in any case. In fact, translation of the 30k, 29k and 23k proteins from their respective purified templates was insensitive to m7GTP in the messenger dependent reticulocyte lysate, as that of the unlabelled partially purified RNA had been (fig. 4), although translation of the 40k-56k proteins was inhibited, suggesting that a contaminating template was responsible for their synthesis. We concluded that these labelled RNA bands were the 30k, 29k (probably) and 23k protein mRNAs respectively, that it is not necessary to 8-eliminate these mRNAs in order to label them with vaccinia capping enzyme preparations and that insensitivity to cap analogue inhibition in the reticulocyte lysate is not simply a reflection of the template activity of a heterogeneous mixture of fragments, but is intrinsic to the purified mRNAs.

While we have found the vaccinia capping enzyme assay, like the m7GTP resistance of template activity in the MDL, to be a useful operational criterion for identifying potential mRNAs and for isolating them in
translatable form, the reaction is not without its difficulties and we are still not certain whether or not the two observations are related as we originally surmised, although this remains a possibility. It appears that the yield of the vaccinia-labellable RNA species is very variable, even after the virus purification procedure was altered to optimise short rod recovery (some of this material is trapped in a green pellicle overlaying the high speed virus pellet which is normally discarded). In all cases recoverable yields of gel purified mRNA are low (in the range of $10^{-2}$ mole% of the total RNA in the preparation shown in Fig. 4). Since yields are low, side reactions catalysed by contaminating enzymes must be considered, especially if their substrate is in excess. In addition, we have found two significantly different patterns of labelled RNA species in different viral RNA preparations, due mainly to RNA species in both kinds of profile which either give no detectable translation products on attempted translation \textit{in vitro} or give a generalised low-level background (for example, the faint band above the 30k mRNA in the gel shown in Fig. 4). The reasons for these observations are not yet clear.

**Detailed Peptide mapping and determination of the N termini of the 29k and 23k proteins.**

The 30k, 29k and 23k protein bands from \textit{in vitro} translation of the separated labelled RNAs in the messenger-dependent reticulocyte lysate were analysed by peptide mapping. $^{35}$S-methionine-labelled maps showed that the proteins were identical to those obtained by translation of unfractionated PSV RNAs, but with less contamination. Both results showed that the sequences of the smaller proteins were subsets of those of the larger proteins (Fig. 5a-c). The 29k and 23k maps, however, each had an unique spot (see below).

Available data (Refs.1,2 and Fig.1) were consistent with a position for the 30k gene between the p126/183 and coat protein genes on the TMV genetic map (Fig. 6). Part of the nucleotide sequence in this region containing an open reading frame was known at the time our experiments were begun since it incorporates the viral assembly origin \cite{33,34}, while the remainder was completed (for both common TMV \cite{4,29} and the closely related Japanese strain OM \cite{35}) while our experiments were in progress, showing that the full extent of the open reading frame is 268 amino acids, consistent with a translation product of about 30k (Fig. 6). The predicted amino acid sequence shows several diagnostic features. First, the number of methionine-containing tryptic peptides expected is 10 — 8 strong spots were found on the 30k map. Second a large tryptic peptide containing three methionine residues spans the
Fig. 5: Peptide maps of $^{35}$S methionine labelled 30k (a,d), 29k (b) and 23k (c,e) proteins translated in the messenger dependent reticulocyte lysate.

Spot numbering and positions of the origin and of dye markers are shown in the key. Cross hatched tryptic peptides are sensitive to V8 protease, open circles denote spots unique to 29k and 23k digests. Spots 1, 5 and 8 are peptides 26, 13 and 4 in Fig. 6, respectively (see text). Of peptides not present in the 23k digest (c), spot 6 is V8 protease sensitive and is therefore peptide 8, while spots 4 and 7 must be peptides 6 and 5 on mobility grounds. Spots 2 and 3 in 23k are probably peptides 18 and 39 respectively, but have closely similar mobilities making identification less certain. Fluorography was used to enhance detection.

centre of the assembly origin while no other peptide contains more than one methionine residue. This peptide is expected to contain 5 glutamic acid residues, and should give rise to 5 peptides (plus free glutamic acid) on digestion with *S. aureus* V8 protease, of which three should be labelled with methionine. Accordingly, peptide maps of 30k and 23k proteins digested first with trypsin and subsequently with V8 protease were examined (Fig. 5d&e). The most strongly labelled peptide (spot 1 in Fig. 6) on the tryptic peptide map was indeed found to disappear, concomitant with the appearance of three other labelled peptides. One additional tryptic peptide was sensitive to V8 protease.
Fig. 6: (a) Open reading frames on TMV RNA and (b) predicted peptides of the Mr 29,987 open reading frame.

In (b) nucleotides are numbered below the sequence starting at the first base of the open reading frame (position 4903 on the RNA). The predicted amino acid sequence is given above the nucleotide sequence using the one-letter code. Trypsin cleavage points (11) and V8 cleavages within tryptic peptides (*) are marked on the line above; tryptic peptides are numbered along the top line starting at the N terminus.

The decreasing messenger sizes for the 30k, 29k and 23k proteins suggested that these proteins might form a C-coterminal set each possessing a different N-terminus. In this case the unique methionine-containing peptides in the maps of the 29k and 23k proteins could be the N-terminal peptides.
retaining the initiating methionines. Considering the sizes of the proteins and the two dimensional mobilities of the unique peptides, examination of the predicted sequence for the 30k protein shows that the most likely AUG initiation codon for the 29k protein is between positions 4960 and 4962, while that for the 23k protein is between positions 5191 and 5193. The corresponding N-terminal tryptic peptides predicted for these sites would be Met-Glu-Lys and Met-Glu-Arg respectively. The unique peptides of the 29k and 23k proteins are both neutral and therefore do not have the correct charge at pH 4.7 to be these peptides. Acetylated forms of Met-Glu-Lys and Met-Glu-Arg, however, would be neutral and have the appropriate chromatographic mobility. To check this interpretation, the unique peptides were eluted and subjected to digestion with carboxypeptidase B. Both peptides gave rise to an identical more acidic methionine-containing peptide, which proved to comigrate with synthetic AcMet-Glu in two dimensions. Given the specificity of carboxypeptidase B for C-terminal arginine and lysine residues, this indicates that the peptides both had the structure AcMet-Glu-Lys\(^\text{Arg}\). Although we did not determine the C-terminal residues directly, we can deduce that the 29k peptide contains lysine and that the 23k peptide contains arginine from the greater mobility of the 23k peptide upon chromatography, since arginine migrates faster than lysine in our chromatography system. Peptides of the mobility expected for Met-Glu-Lys and Met-Glu-Arg were present in the map of the 30k protein. The map of the 29k protein contained only the putative Met-Glu-Arg spot, while that of the 23k protein had neither. These spots were eluted and treated with carboxypeptidase B. Both gave rise to an identical more acidic peptide which comigrated with Met-Glu. Thus spot 8 is Met-Glu-Lys and spot 5, which migrates slightly faster upon chromatography, is Met-Glu-Arg. Note that none of the four tripeptides is susceptible to digestion with V8 protease. We suspect this may be due to an inability of V8 protease to act as a carboxypeptidase, perhaps compounded by a neutralizing effect of the basic residue adjacent to the glutamic acid.

All the above results are consistent with the identification of the methionine codon at positions 4960-4962 as the 29k protein initiation site and the methionine codon at positions 5191-5193 as the 23k protein initiation site. Translation of the 30k open reading frame from these sites to the ochre termination codon would give proteins of 249 and 172 amino acids, corresponding to Mr's of 27,875 and 19,478 respectively. As expected all the methionine-labelled tryptic peptides of the 30k protein are present in the map of the 29k protein making allowance for the mobility shift caused by acetyl-
ation of Met-Glu-Lys. The 23k protein map, however, contains only four out of eight peptides. This is exactly as predicted by the sequence, since four methionine-containing tryptic peptides occur between positions 4960 and 5191. One of the remaining four tryptic peptides is expected from its mobility, given the sequence, to be the penultimate tryptic peptide of the 30k open reading frame (see legend to Fig. 5). We conclude that the 29k and 23k proteins are initiated within the 30k open reading frame and are almost certainly C-terminally with the 30k protein. No unique peptide was detected corresponding to the expected N-terminal methionine of the 30k protein. It seems most likely that the initiating methionine is removed from this protein post-translationally. In addition, one V8 protease sensitive spot is predicted to occur in the 23k protein, corresponding to the very hydrophobic peptide 21. There is a candidate for this spot, running above spot 1, but its yield is low and variable and we have not included it either in the above tally or in the key to Fig. 5 where it is only faintly visible. With these exceptions every predicted methionine-containing tryptic peptide on the maps is clearly visible on the maps and can be accounted for in a manner consistent with the sequence and the proposed mode of overlapping (compare Figs. 5 & 6). The protein molecular weights predicted are in fair agreement with the observed values for the 30k and 29k proteins, but the 23k protein has a calculated Mr of only 19,500 — apparently its migration on SDS gels is anomalously slow. According to the convention outlined earlier, based on their true molecular weights these three proteins should now be designated p30, p28 and p19, respectively, although for the sake of uniformity we shall continue to use the old nomenclature throughout the remainder of this paper.

DISCUSSION

This work began with the aim of using PSV-RNA translation to locate the 30k gene more precisely on the TMV genome and peptide mapping to distinguish the 30k protein from the other known TMV proteins. However, it soon became obvious that we were dealing not with one protein, but a family of related polypeptides with distinct mRNA templates (Fig. 3). Since then, we have been mainly concerned with two interrelated questions: first, how are the proteins and their messengers related, and second, can we be sure that the appearance of so many related proteins is not an artifact? We have not attempted a definitive catalogue of all the proteins and mRNAs. Data reported in this paper define the structural relationship between the major 30k-related proteins. The 30k, 29k and 23k proteins form a nested set with
staggered N termini and probably identical C termini. The latter conclusion is supported by the fact that antibodies directed against a synthetic peptide corresponding to the C-terminal 11 amino acids of the 30k open reading frame precipitate all three in vitro products, plus a number of others including three corresponding in size to the remaining proteins seen in Fig. 3. Of these the 28k protein has been shown by peptide mapping to conform to the established overlapping pattern, while an mRNA has been found for the 18.5k protein using the vaccinia capping enzyme reaction, extending the correlation established for the major members of the set (our unpublished results).

Obviously such a multiplicity of proteins could be artifactual. But equally obviously, "early quitting" (premature termination of translation) cannot explain the data, since the proteins differ at the wrong end. This leaves us with the possibility of artifactual synthesis on templates of fragmented RNA. In that case, since different sizes of RNA direct synthesis of correspondingly sized proteins with differing N termini, one must postulate preferential breakage of genomic RNA at various points within the 30k gene, in each case producing an RNA fragment extending from the break to the genomic 3' terminus (to account for the template sizes) on which the first AUG from the break is unmasked as an initiation codon. Such fragments would be hard to distinguish from subgenomic RNAs. However, the major 3' coterminal fragments produced deliberately by alkaline stripping and nuclease digestion of common strain TMV are inefficient templates for synthesis of the 30k and related proteins (Fig. 3). The bulk of the sucrose gradient fractionated short RNA extracted from native virus particles is also separated from the vaccinia labelled mRNAs on 3% urea gels, as judged by ethidium bromide staining (data not shown). These minor RNAs direct efficient synthesis of at least three of the 30k-related proteins (Fig. 4), and in at least two cases preferentially direct the synthesis of a single protein. Both these observations are consistent with, although they certainly do not prove, the hypothesis that these RNAs are functionally monocistronic mRNAs made in vivo which appear in the population of packaged RNAs because they contain the viral assembly origin, which forms part of the structural gene for all three proteins.

Against this evidence we have to set the anomalous responses of the packaged mRNAs to cap analogues, which might suggest in the case of the smaller 30k related proteins that their synthesis is an artifact of RNA degradation (24). However, the 30k mRNA activity itself exhibits the same anomalous responses to cap analogues and nuclease treatment as its smaller relatives, and is clearly distinguishable on both counts from coat protein
mRNA. If the template activities for the smaller proteins are judged spurious on that basis, it follows that the 30k mRNA activity cannot be physiological either, despite representing complete translation of the relevant open reading frame on TMV RNA\(^{(4,35)}\), which is known to be expressed from genetic evidence (Ref. 36 and unpublished results of DZ and TH). Therefore this interpretation leads to the conclusion that either the 30k mRNA is not packaged at all, or is packaged in such small amounts as to be swamped by background synthesis from fragmented RNA. Although our data cannot exclude this possibility, the idea clearly raises as many problems as it solves. Northern blotting experiments using RNA from infected leaves\(^{(37,38)}\) indicate that most if not all the intracellular TMV RNA species larger than the coat protein mRNA are packaged. In fact, the large number of p30-related RNAs that we observe is quite consistent with the many TMV-related RNAs observed by Northern blotting, although RNAs appearing on our gels represent a size selected sample excluding the larger species seen in those studies, and possibly resolving more species in a small part of the molecular weight range. Though Northern blotting has the disadvantage that hybridised RNA cannot be directly recovered from the filter for translation, while controlling for RNA breakdown is still a problem, both approaches suggest that there may be more subgenomic RNAs than there are open reading frames on TMV RNA.

On balance, it seems more likely that at least the major 30k related mRNA activities observed in vitro are genuine, if unusual, since the evidence to the contrary is certainly not conclusive. Thus, the apparent activation of these mRNA activities by mild nuclease digestion may simply be due to their much smaller target size compared to the templates for pl26 and pl83. A possible explanation for their anomalous behaviour towards cap analogue inhibition is suggested by the observation that vaccinia capping enzymes will label the putative mRNAs without 5'-elimination. As we have stressed, some reservations are necessary but the simplest interpretation is that these RNAs (or a proportion of them) are not capped, but di- or tri-phosphorylated, which in turn implies that they carry an intact 5' end. (Since the quantity of RNA in a labelled band is below the level of detection of ethidium bromide staining we cannot calculate exactly what percentage of the substrate RNAs are uncapped.) Satellite tobacco necrosis virus RNA is known to have a di- or tri-phosphorylated 5' end\(^{(39,40)}\), and its translation resembles that of the 30k family of mRNAs in responding in a complex way to cap analogues\(^{(26)}\). Phosphorylated, uncapped termini might also account for the different responses of translation of the 30k-related proteins to m^7 CTP and SAH in the
wheat germ and reticulocyte systems, since the wheat germ system contains capping enzymes which would cap and methylate di- or tri-phosphate 5' ends whereas the reticulocyte lysate does not\(^{(27)}\). The main difficulty with this idea is that we have to postulate two genes on the same virus, both expressed via subgenomic RNAs, yet one mRNA being capped and the other group of mRNAs not. So far though the data, while indirect, appear to point in this direction. In order to test this model, experiments are in progress to see whether the proteins identified here as products of in vitro translation can be identified in vivo using the antibodies we have raised against a synthetic 30k peptide.

ACKNOWLEDGEMENTS

We thank Christine Harley for technical assistance, Tim Hunt for performing an experiment on the evening of a May Ball, Hugh Pelham for discussions, Bernard Moss, Milt Zaitlin and Jo Butler for gifts of materials and Philip Goelet for providing Fig. 6. Special thanks to Philip Goelet and Yoshimi Okada for access to their and their colleagues unpublished RNA sequences.

REFERENCES