Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Region of Tobacco Mosaic Virus RNA Involved in the Nucleation of Assembly

D. Zimmern

Phil. Trans. R. Soc. Lond. B 1976 **276**, 189-204
doi: 10.1098/rstb.1976.0110

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1976 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Phil. Trans. R. Soc. Lond. B. **276**, 189–204 (1976) [189]
Printed in Great Britain

The region of tobacco mosaic virus RNA involved in the nucleation of assembly

By D. Zimmern
M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH

[Plates 29 and 30]

The interaction of TMV RNA with the disk aggregate of TMV protein at the initiation of assembly has been studied by using the techniques of RNA sequencing. The 5′ end group has been identified, and shown not to be protected in the early stages of assembly from accessibility to nuclease digestion. A population of RNA fragments of average length 250 nucleotides, originating from a unique region of TMV RNA, is encapsidated by limited assembly, and sufficient sequence information is available to identify certain unusual features. The protected region does not contain highly reiterated simple repeating sequences, but may contain more complicated repeats. The length and complexity of the nucleation region may reflect adaptation to the efficient mediation of the conformational change from disk to helix of TMV protein, besides a requirement for binding to the disk, and this may be an important part of the mechanism of specificity in the nucleation of assembly.

1. Introduction

Fraenkel-Conrat & Williams demonstrated the reassembly of infectious virus-like particles from separated TMV RNA and coat protein in 1955 – the first demonstration of a simple self-assembling nucleoprotein complex and the prototype virus self-assembly system. Since that time, further work has shown that the protein has a special adaptation relevant to assembly, namely that it can exist as a disk-shaped aggregate consisting of two layers each of 17 subunits (Durham 1972). This aggregate is required for initiation of assembly (Butler & Klug 1971) as has been confirmed by others despite disagreement about the role of the disk in elongation (Richards & Williams 1972; Okada & Ohno 1972). In its 17-fold rotational symmetry the disk approaches that of the virus which has $16\frac{1}{3}$-fold rotational symmetry, and it seems reasonable to think of it acting at this stage as a sort of jig.

It has been felt for some time (e.g. Caspar 1963) that the RNA would also prove to contain some region specially adapted to the needs of assembly, since the reaction shows a marked preference, though not an absolute requirement, for the homologous RNA. Experiments designed to test this hypothesis led to the conclusion that a short region of the RNA within 50 nucleotides at most of the 5′ terminus was involved (Butler & Klug 1971; Thouvenel, Guilley, Stussi & Hirth 1971; Guilley, Stussi & Hirth 1971; Ohno, Nozu & Okada 1971). The fact that poly A could also be encapsidated by TMV protein (Fraenkel-Conrat & Singer 1964), albeit by a mechanism kinetically different to that with the viral RNA (Butler 1972), led to the speculation that an adenine-rich region was involved. The 5′ terminal nucleotide itself was thought to be an adenine residue which did not carry the 5′ terminal di- or tri-phosphate group characteristic of most primary transcripts (Fraenkel-Conrat & Singer 1962; Sugiyama & Fraenkel-Conrat 1963; Fraenkel-Conrat & Fowlks 1972). Since binding of RNA to the disk was expected to be a highly cooperative process, it was widely thought that the relevant sequence might

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.* www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

contain heavily repeated features, probably with a repeating period of three nucleotides, since each coat protein subunit binds three nucleotides in the structure of the mature virus (Caspar 1963). Taken to their logical conclusions, these models collectively predicted a sequence of up to seventeen identical or substantially similar adenine-rich triplets sequentially arranged at the extreme 5′ terminus of TMV RNA. More than three consecutive repeats of the triplet AAA itself were excluded since analysis of large oligonucleotides produced by digestion of TMV RNA with RNase T_1 showed no larger oligo-A sequence, as is evident from, though not explicitly stated in, the work of Garfin (Garfin 1972; Garfin & Mandeles 1975). It would in any case be difficult to see how poly-A and TMV RNA could initiate assembly with different orders of reaction if this were the only feature important in initiation of assembly. It is also worth noting at this stage that while heavily repeated parts of the sequence might facilitate cooperative and therefore tight binding to a polymer of identical subunits such as the disk, such reasonably tight binding of the RNA may be a necessary but not *of itself* a sufficient condition for nucleation of assembly. The unique role of the nucleation region of TMV RNA in assembly is to direct the conformational change ('dislocation') of *one* disk to give a step at one place, without having the added cooperative interactions with the already helical growing tip of a partially assembled nucleoprotein rod to help supply the necessary activation energy. (The evidence for, and a model of, the process of initiation via dislocation of a single disk is given in an earlier paper in this volume by Butler.)

The nucleotide sequence requirements for binding to and for efficiently dislocating the disk may well be different, indeed the aberrant kinetics of initiation of assembly with poly A suggests that they are. Furthermore, apart from possible sequence specific interactions at nucleation of assembly, TMV protein must bind *all* the RNA in the structure of the mature virus, and relatively sequence-unspecific nucleotide binding must therefore of necessity be a feature of the molecule in the conformation found in helical aggregates. It is noteworthy that if fragments of TMV RNA are bound to the disk, various lines of evidence suggest that aberrant particles are formed which incorporate RNA not normally recognized at initiation (Guilley *et al.* this volume; Tyulkina *et al.* 1975). This may reflect both the lability of the RNA region involved and, more importantly, that the possibility exists for relatively unspecific nucleotide binding by the protein even in the disk state. Perhaps this mode of RNA binding is related to the normal mode, and the sequences selected related to the sequence preferentially recognized, but there must obviously be some reason why these sequences are not normally recognized, which may be a result of a deficiency in steps *subsequent* to binding. This point will be considered further in a later section.

Sequence analysis of the nucleation region of TMV RNA was initiated in our laboratory three years ago at a time when it was thought that nucleation with TMV RNA, as with poly A, might involve two disks sandwiching the crucial piece of RNA. The experiments were designed with this outline description of the process in mind, in the hope of filling in a few of the exact molecular details. However, the results indicate that this currently accepted description of the RNA and of its expected role in assembly requires substantial revision to account for the facts.

We have used two different methods to study the RNA sequence involved in nucleation. The first and simplest involves using the specific labelling of 5′ termini of polynucleotides with ^{32}P using the enzyme polynucleotide kinase (Richardson 1965). The second and potentially most informative method involves protection of totally ^{32}P labelled TMV RNA from nuclease

Downloaded from rstb.royalsocietypublishing.org

digestion with limiting amounts of coat protein, followed by re-extraction of the protected RNA and sequence analysis. This is an application of a protocol first used to sequence ribosome binding sites from bacteriophage RNA (Steitz 1969; Hindley & Staples 1969).

2. Polynucleotide kinase and the 5′ end group of TMV RNA

The polynucleotide kinase catalysed phosphorylation of the 5′ end of TMV RNA has previously been studied by two groups of workers (Suzuki & Haselkorn 1968; Fraenkel-Conrat & Fowlks 1972). The former group concluded that TMV RNA was a very poor substrate for the enzyme, accepting only 0.3 mol of phosphate per mole of RNA. The reason for this was unexplained, but these workers suggested that the 5′ terminus was unavailable, either because of being folded away behind some extensive secondary structure, or because of the presence of some form of blocking group. It is not clear from the published data whether a simple phosphorylated terminus had been excluded in this case as the blocking group by pretreatment with a phosphatase to remove the 5′ phosphate(s). Such a control was performed with TYMV RNA, which was studied at the same time, and for which phosphatase treatment before labelling made no difference. The latter group succeeded in raising the amount of label incorporated to near stoichiometric amounts, but although analysis then showed pAp to comprise about 50% of labelled termini, all three other nucleotides were also labelled, and moreover fingerprints of the labelled material were 'heterogeneous'. This was interpreted as being due to heterogeneity of the 5′ end itself.

We attempted similar experiments, using conditions substantially similar to those of Suzuki & Haselkorn, though scaled down and with [^{32}P]ATP of higher specific activity to allow application of the fingerprinting techniques of Sanger and co-workers, which are limited in the amount of material that can be fractionated. Apparently full length 5′ labelled TMV RNA was isolated by sucrose gradient centrifugation, and fingerprinted. The stoichiometry of incorporation of phosphate label was about 0.2–0.3 mol per mole of RNA, and furthermore this label was distributed seemingly at random among the products of digestion with either of RNases T_1 or A. The pattern of these products was unaffected by pretreatment with *E. coli* alkaline phosphatase, which increased the molar amount of label only slightly. Heating to 80 °C for 5 min before alkaline phosphatase treatment and labelling also made very little difference to the amount of incorporation, although this did not rigorously exclude steric hindrance of access by the enzyme to the 5′ end, since renaturation could be very rapid after cooling.

The complexity of these fingerprints (figure 1, plate 29) was such that labelling of a 5′ terminus that was itself heterogeneous could be excluded on these grounds alone, as a substantial fraction of the entire sequence complexity of TMV RNA is represented in the patterns. It seemed more reasonable to suppose that incorporation was occurring at nicks in the RNA, which would still sediment in the same position on sucrose gradients provided the molecule was held together by non-covalent forces. In support of this, electron microscopy of apparently homogeneous 30S (i.e. full length) TMV RNA molecules purified by sucrose gradient sedimentation showed about 25% broken molecules when examined under denaturing conditions (B. Jacrot, personal communication). Since this correlated well with the molarity of incorporation of phosphate, the enzyme was obviously working and the question arose as to what was preventing incorporation at the true 5′ end.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

A possible answer was suggested by reports originating in the work of Perry & Kelley (1974) that eukaryotic messenger RNAs were methylated. Some features of the methylated nucleotides implicated the 5′ terminus as their origin – for example the high salt concentration required to elute them from DEAE-cellulose columns, and the observation by Furuichi (1974) that transcription of mRNA from the double stranded RNA genome of CPV (cytoplasmic

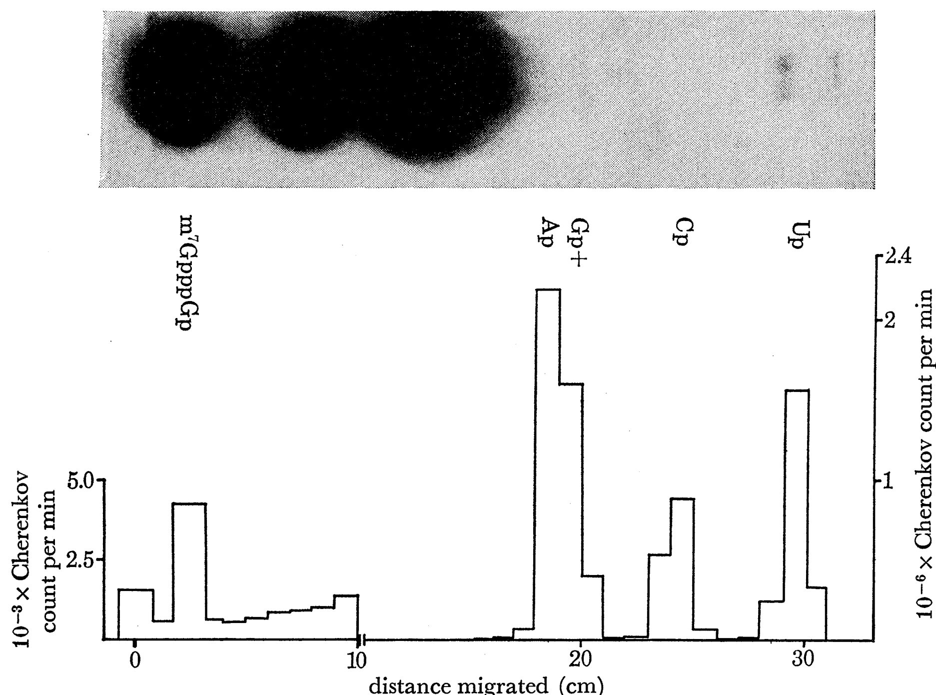


FIGURE 2. TMV RNA contains an oligonucleotide resistant to RNase T_2, of structure $m^7GpppGp$. An RNase T_2 digest was applied to Whatman DE 81 paper and electrophoresed at pH 3.5 (electrophoresis from left to right).

polyhedrosis virus of silkworms) was coupled to and dependent on methylation, the methyl groups being incorporated into very short nascent chains. Incompletely characterized methylated 5′ termini had already been demonstrated in the sense strand of CPV genome RNA (Miura, Watanabe & Sugiura 1974) and in mammalian low molecular mass nuclear RNA (Reddy, Ro-Choi, Henny & Busch 1974). We wondered if some such modification, perhaps a ribose methylation, at the 5′ end of TMV RNA was preventing efficient recognition of the terminus by polynucleotide kinase. We therefore screened complete RNase T_2 digests of totally ^{32}P labelled TMV RNA for the presence of an unusual nucleotide, and found among the products of digestion a nucleotide that migrated very slowly compared to the major mononucleotide

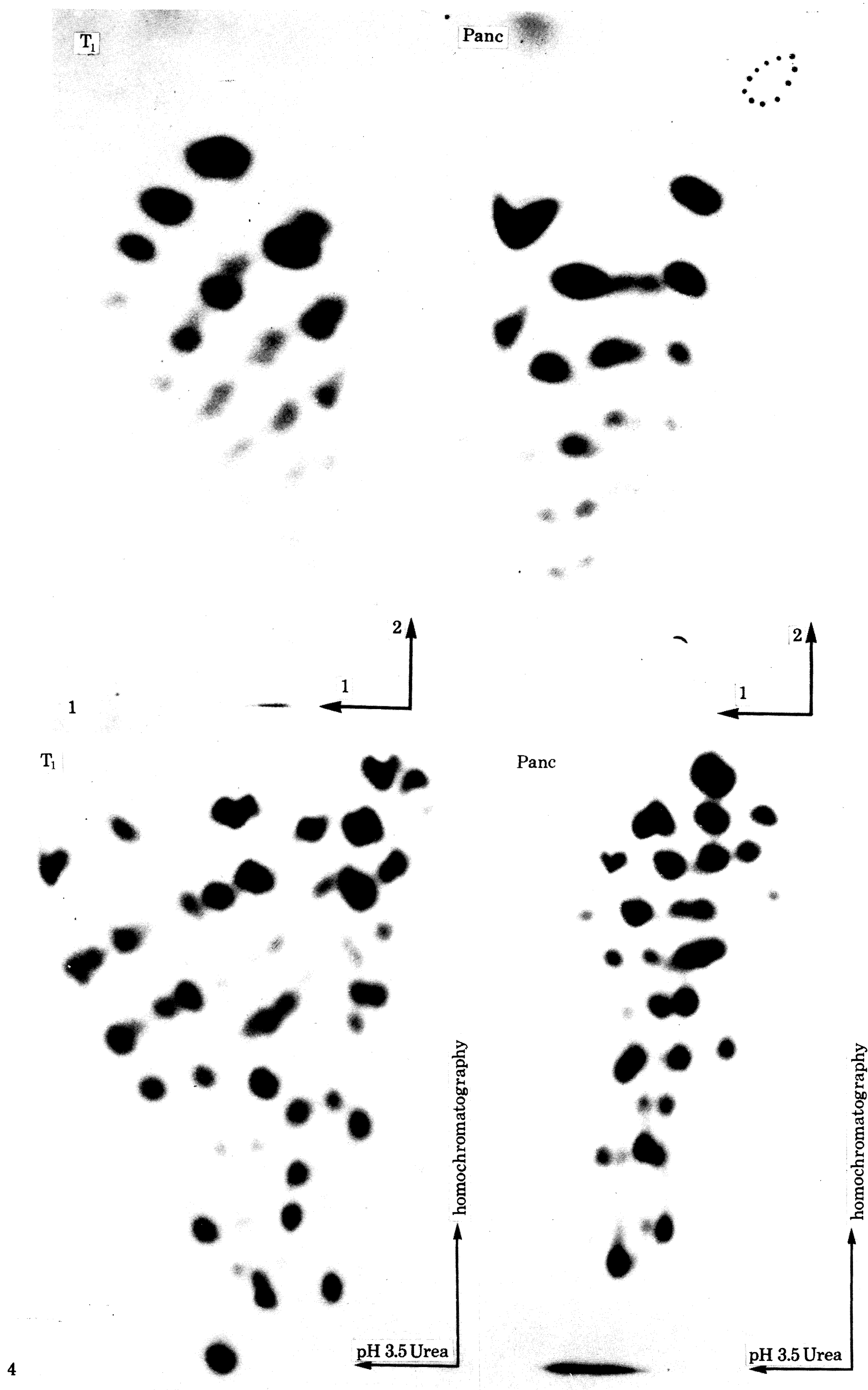
---

DESCRIPTION OF PLATE 29

FIGURE 1. T_1 and pancreatic ribonuclease fingerprints of polynucleotide kinase labelled TMV RNA, purified by sucrose gradient centrifugation. Left panel, T_1 RNase fingerprint. Right panel, pancreatic RNase fingerprint. 1st dimension: electrophoresis at pH 3.5 on cellulose acetate in pyridine acetate buffer containing 7 M urea and 5 mM EDTA. 2nd dimension: homochromatography, 3 % 20 min hydrolysed homomixture.

FIGURE 4. T_1 and pancreatic RNase fingerprints of RNA extracted from complexes prepared by limited assembly. Left panel: T_1 RNase fingerprint. Right panel: pancreatic RNase fingerprint. Conditions as in legend to figure 1.

Downloaded from rstb.royalsocietypublishing.org



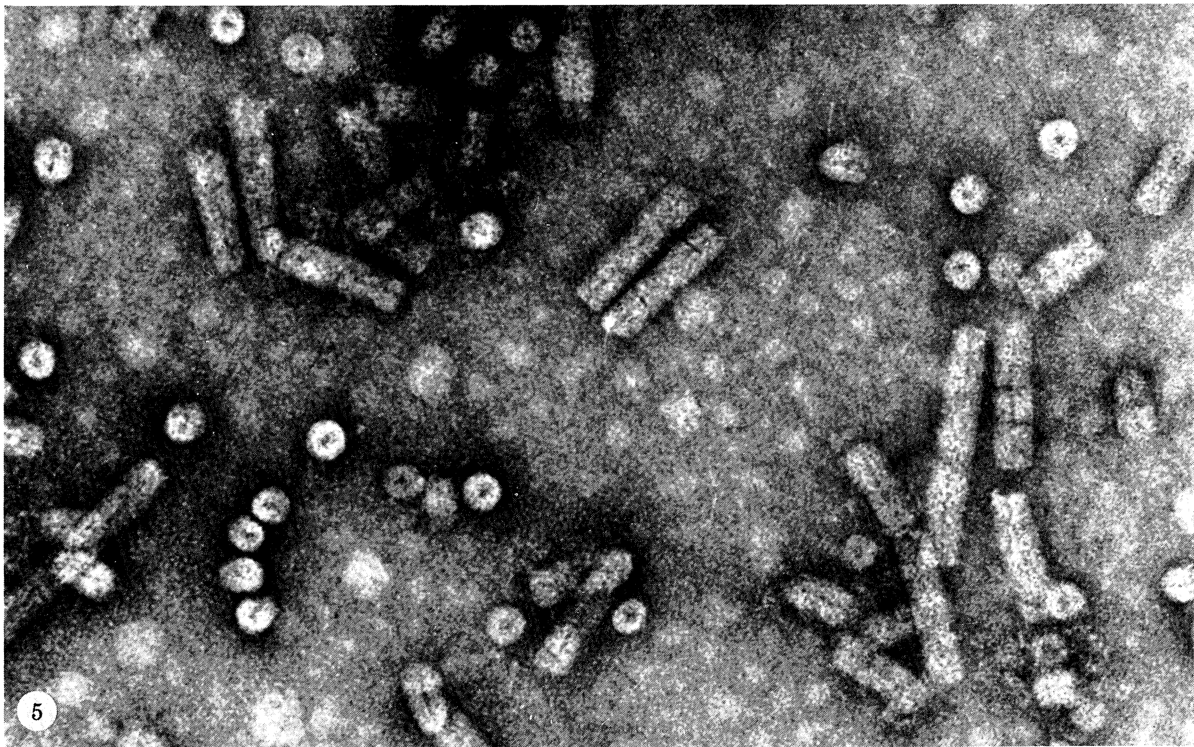
Figures 1 and 4. For description see opposite.

(*Facing p.* 192)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



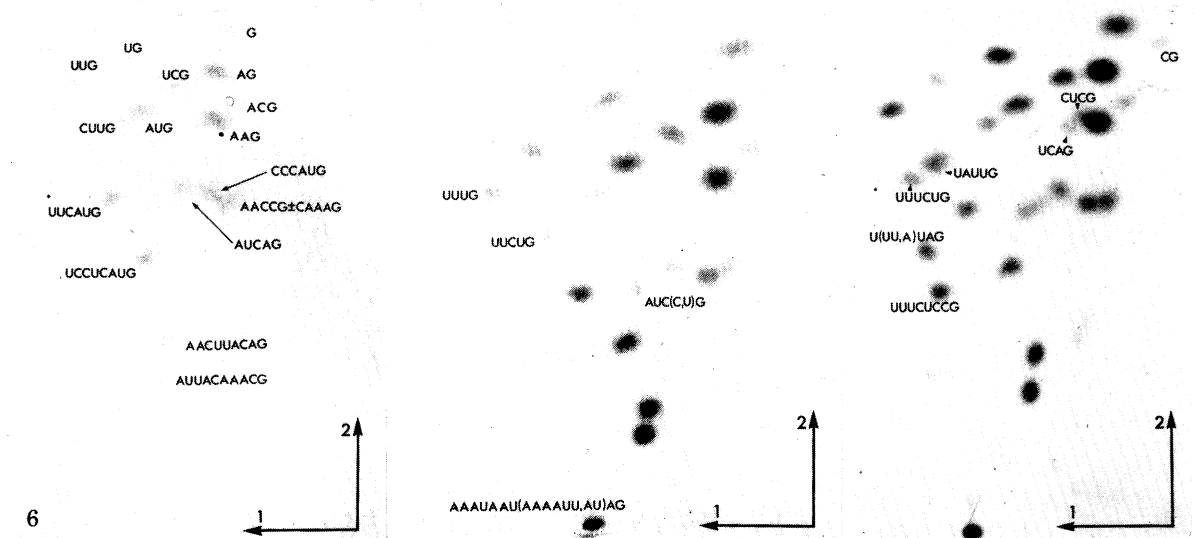


FIGURE 5. Appearance in the electron microscope of material resulting from limited assembly followed by nuclease treatment. (Magnification approximately × 400 000.)

FIGURE 6. T_1 RNase fingerprints of bands eluted from a polyacrylamide-urea gel used to fractionate the protected RNA population. Sequences of spots are given when they first appear.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

products when analysed by electrophoresis on DEAE paper (figure 2). Such behaviour is characteristic of oligonucleotides or acidic nucleotides arising respectively from ribose methylation (which prevents cleavage by RNase T_2) or possession of acidic substituents, most commonly extra phosphate groups in nucleotides arising from the 5′ end of RNA molecules. A similar procedure was originally used to screen partial digestion products of R17 RNA for the presence of 5′ terminal pppGp (Adams, Spahr & Cory 1972). The finding of a nucleotide moving with the electrophoretic mobility of a phosphorylated end group in limit digests of TMV RNA was contrary to previous reports (Fraenkel-Conrat & Singer 1962; Sugiyama & Fraenkel-Conrat 1963). Further analysis showed that the structure of this minor component had a similar configuration to the blocked, methylated 5′ terminal structures that have by now been found in mRNA from three groups of animal viruses (Furuichi & Miura 1975; Furuichi, Morgan, Muthukrishnan & Shatkin 1975; Wei & Moss 1975; Urushibahara, Furuichi, Nishimura & Miura 1975; Abraham, Rhodes & Banerjee 1975), one other group of plant viruses (Kaesberg 1975) and from uninfected mammalian tissue culture cells (Adams & Cory 1975; Wei, Gershowitz & Moss 1975; Furuichi *et al.* 1975). In this case the 5′ terminal structure is m^7GpppGp, where the m^7G is inverted with respect to the rest of the RNA chain and forms a blocking group, which prevents access by polynucleotide kinase to the 5′ hydroxyl group, while phosphatase treatment makes no difference to the accessibility of the 5′ end. The detailed evidence for this structure is presented elsewhere (Zimmern 1975).

As far as the assembly of TMV is concerned, the main consequence of the presence of this blocking group at the 5′ end is that it removes most of the evidence for a 5′ to 3′ polarity for assembly. A polar mechanism starting at most within 50 nucleotides (i.e. one turn of virus helix) from the 5′ end had been inferred from two sorts of experiment. The first involved digestion of TMV RNA with exonucleases. It was shown that assembly was inhibited by preparations of spleen phosphodiesterase, a 5′ to 3′ exonuclease, with the accompanying release of between 15 (Guilley *et al.* 1971) and 50 (Butler & Klug 1971) nucleotides. Venom phosphodiesterase, a 3′ to 5′ exonuclease, had no effect (Butler & Klug 1971). However, since spleen phosphodiesterase requires a free 5′ hydroxyl group, and will not digest nucleotides blocked with m^7G (Reddy *et al.* 1974; Wei & Moss 1975), this inhibition is clearly not due to exonuclease action. It seems most probable that a contaminating endonuclease is responsible, as the digestion leads to breakdown of the RNA with a resulting decrease in sedimentation coefficient besides release of acid-soluble material, although such degradation is not over-extensive at the point where assembly activity is lost. Therefore, if inhibition of assembly is due to the action of a contaminating endonuclease, the relevant sensitive part of the molecule must be among the more labile to the action of that endonuclease. The most probable candidate for the endonuclease is spleen acid ribonuclease (Bernardi & Bernardi 1968) which, in the most extensively documented case (that of *E. coli* 5S RNA), was found to cleave preferentially at A–U and G–U bonds when used to generate partial digestion products (Brownlee 1972). However, it is not clear that one of the vulnerable sites of cleavage actually forms part of the disk recognition site, since Guilley *et al.* (1971) noted a large number of very short rods of modal length 38 nm (800 nucleotides of RNA) in reconstitution mixtures with spleen phosphodiesterase-treated RNA. It could therefore be the case that the site is excised intact within this 800-nucleotide segment.

The second sort of experiment from which a 5′ to 3′ polarity was deduced involves several variations on the theme of periodate oxidation of the free 2′,3′ *cis*-hydroxyl group at the 3′ end

Downloaded from rstb.royalsocietypublishing.org

of the RNA, followed by labelling of the carbonyl groups so generated. This can be used as a probe of assembly either by attempting to label RNA fragments protected by partial assembly from nuclease digestion, or by seeing whether label is released from terminally labelled RNA by nuclease after partial assembly, or by other related experiments (Thouvenel *et al.* 1971; Ohno *et al.* 1971). Although these experiments are slightly less direct than the first sort (most not distinguishing, for example, between straightforward 5′ to 3′ polarity, and a mechanism involving initiation near the 3′ end, a fast 3′ to 5′ growth and a slow annealing to protect the 3′ end group), the results agree that periodate-sensitive groups are not encapsidated early in assembly. The snag in interpretation that results from the blocking group's presence at the 5′ end is that this also possesses a free 2′,3′ *cis*-hydroxyl group because of its inverted orientation with respect to the rest of the chain, and therefore it too should be periodate sensitive. On the face of it, therefore, these results must now be interpreted to mean that *neither* end group is encapsidated early in assembly. (This interpretation assumes that both end groups are equally sensitive to reaction with periodate. It should also be noted that Ohno *et al.* (1971) found about 25% of the input label in partially reassembled material when terminally pre-labelled RNA fragments were used for assembly (their figure 3*b*).)

Apart from the chemical evidence, the best remaining support for a polar model comes from observations of partially assembled rods in the electron microscope (Stussi, Lebeurier & Hirth 1969; Rodionova, Vesenina, Kichatova & Atabekov 1971; Butler & Klug 1971). These are usually found to have only one RNA 'tail' even when quite short rods are examined. A few rods may be seen on some plates, especially those made by spray-droplet techniques and shadowing, which have two tails. However these are a minority and care is needed in interpreting their significance since the spray-droplet technique may result in loss of protein from the tips of the rods during sample preparation (see earlier paper by Butler, this volume). A conservative estimate of the minimum size of rod to show further growth from one end only puts the site where assembly initiates within 1000 nucleotides of one end. We have evidence (D. Zimmern, T. Hunter, T. M. A. Wilson, R. N. Perham & B. G. Barrell, unpublished), that the 3′ terminal 800 nucleotides contain the coat protein cistron, and have no large T_1 nucleotides in common with the first 300 to be encapsidated (next section). Therefore initiation of assembly probably occurs within 1000 nucleotides of the 5′ end, and for most of the length of the virion (about 5000 nucleotides out of 6400) is polar. However, these observations cast no light on the initial stages of assembly, particularly since a short second tail would probably not be resolved in the electron microscope. All available chemical evidence suggests that neither end group is protected by coat protein early in assembly, and unless these results are artefactual it is necessary to postulate such a short second tail. Obviously there are conceptual difficulties in a model involving a polar structure assembling in two directions, which have so far led to such interpretations being avoided. It remains to be seen whether the more complicated model is supported by further experiments, or whether the experiments so far are misleading and the true explanation is more simple. However, assuming that the conclusions are correct, there are basically two ways in which the second tail could be protected; either by a slow annealing process utilising coat protein in the opposite direction to the major assembly process, or else by the binding of another sort of protein entirely, i.e. a maturation protein. Although such a protein has not so far been found in TMV preparations the expected 2000-fold molar excess of coat protein in mature TMV forms a background against which a minor protein might so far have escaped detection.

Downloaded from rstb.royalsocietypublishing.org

3. The RNA sequence coated early in assembly

(a) *Basic features*

Protection of nucleotide sequences from nuclease digestion by the binding of proteins that interact with them has proved extremely useful for the isolation of diverse functionally important regions of DNA and RNA molecules. When this approach is attempted as a way of operationally defining and isolating the region of TMV RNA which is first coated by limiting amounts of the disk aggregate of the coat protein, a small percentage of the totally

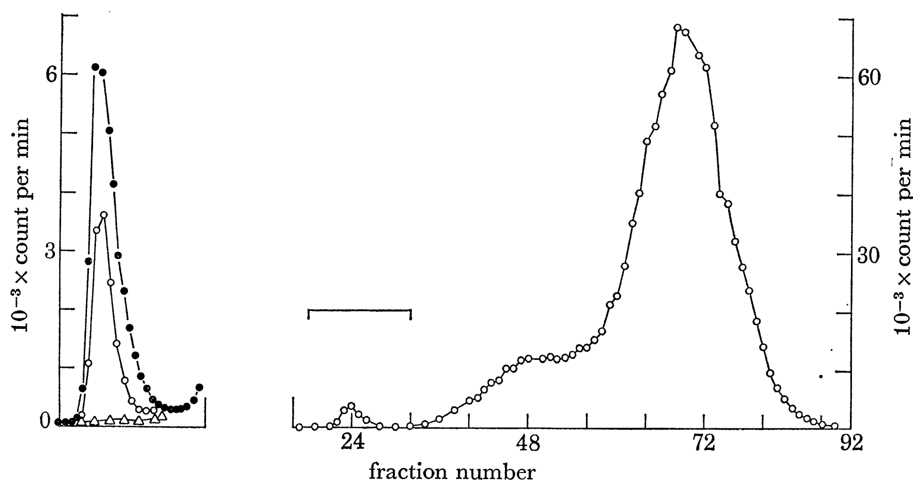


Figure 3. Isolation of T_1 RNase resistant nucleoprotein complexes resulting from limited assembly by gel filtration on Sephadex G-200. ●, 2 disks/RNA; ○, 1 disk/RNA; △, 4S protein.

labelled RNA is protected (this was first shown by P. J. G. Butler, unpublished results). The nucleoprotein complex may conveniently be freed from nuclease and digested RNA fragments by gel filtration on Sephadex G 200 (figure 3). The protected nucleoprotein appears as a small peak of counts running in the void volume. This excluded material is missing if the incubation is performed starting with disaggregated 'A' protein under otherwise identical conditions, showing that disks are necessary for this reaction as for initiation of assembly. The protected RNA gives characteristic and extremely reproducible fingerprints after re-extraction from the complex (figure 4, plate 29) showing that a unique region of the RNA is being encapsidated. Several details deserve comment.

First, both end groups are missing from the fingerprint, and detailed sequence analysis shows that both ends of the protected RNA are frayed. As discussed above, this agrees with other evidence in suggesting that the initiation complex is a two-tailed structure before digestion. A second tail need not be entirely free – it could lie on the surface of the protein providing that it was accessible to nuclease, at least at this stage. Thus it could be only one turn of helix (50 nucleotides) long, requiring only a 'finishing-off' addition of protein subunits around that turn, which would avoid the need to postulate extensive growth in that direction. It should be emphasized, however, that there is no evidence that this second tail is not actually much longer.

If the lack of one or other end group does not reflect the true situation before digestion then it is necessary to postulate some sort of protein redistribution or selective loss of protein from the end. It is not immediately obvious *a priori* why protein should be displaced from the end of the

Downloaded from rstb.royalsocietypublishing.org

RNA by nuclease digestion if that end of the RNA was completely protected in the first place. Moreover the fingerprint of the protected RNA is invariant over a range of digestion conditions that differ in enzyme concentration and time of digestion by more than 1 order of magnitude, therefore any protein displaced must be much more weakly bound than the remainder. RNA rebinding experiments (see below) also suggest that even if the end were initially protected and then lost it could not have been essential to disk binding, since the RNA which is protected seems fully competent to rebind to the disk.

Secondly, the fingerprint is rather complicated, certainly more so than would be expected if the added protein were distributed uniformly over the population of RNA molecules. Polyacrylamide gel electrophoresis of the protected RNA (made under conditions where uniform addition would completely coat only 150 nucleotides, possibly with 50 at each end partially protected) shows quite strong bands of up to 450 nucleotides chain length, although analysis of oligonucleotide yields from fingerprints of the unfractionated material suggests most rods are protecting 250–300 nucleotides of RNA. From a technical point of view, this means that with the amount of radioactive TMV RNA conveniently available (approximately one order of magnitude lower than with bacteriophage RNA) sequencing of an RNA fragment of this size approaches the limits of classical techniques. A substantial amount of the sequence has however been deduced. From the point of view of assembly, on the other hand, the size of the protected region implies that growth as well as initiation has occurred, and that nucleation alone cannot be obtained by this method. In fact, the addition of more protein to an already nucleated rod seems favoured over the nucleation of further rods to a quite surprising extent in conditions of very low protein input. Thus the percentage of input ^{32}P in the RNA protected in a typical experiment is approximately 2.4 ± 0.3 % from an input of protein equivalent to two disks per RNA molecule, which is theoretically capable of protecting 3.2 % of the RNA if the structures formed have the same RNA: protein ratio as the virus (in fact they have slightly less RNA – see below). If only one disk's worth of protein is used the proportion of protected radioactivity drops to about half this (see scale expanded panel, figure 3), however this material when re-extracted and fingerprinted gives the same fingerprint as the material protected by two disks. In this latter experiment, given the chain length of the protected fragment, one can calculate that more than half the input RNA does not nucleate assembly at all.

Since extension to a length of at least 200 nucleotides seems to be so greatly favoured, one may postulate that coating of at least this much of the sequence corresponds to classical nucleation despite the fact that it is four times longer than could be bound to one disk in a single turn. Unfortunately, since there is no way to restrict TMV assembly to the disk-dependent initiation step, in the way that ribosome binding restricts translation to an initiation step, it is not possible to decide directly whether this is so, or whether it is simply a consequence of slow nucleation followed by fast growth from the disk preparation. If one looks for clues in the sequence of the protected RNA the most unusual part, in terms of clusters of like residues, occurs in the region from residues 50–200, approximately (§3*c*). On the other hand, bands can be resolved on polyacrylamide gels from among the population of protected fragments that correspond to the first hundred nucleotides only, i.e. to the addition of a single disk. Is this latter observation incompatible with the involvement of the more distant parts of the sequence in the mechanism of packaging? In other words, must unusual features not within the first 50–100 nucleotides be dismissed as irrelevant to nucleation? Some apparent contradictions here result, I think, from the application of mechanistic models which think of the

Downloaded from rstb.royalsocietypublishing.org

RNA and protein components of the reaction more or less as black boxes. One needs to bear the following points in mind:

(i) As discussed in the introduction, there is reason to believe that binding of the RNA to the disk may be a necessary but not sufficient condition for nucleation of assembly. If one part of the sequence were adapted for efficient binding of the disk and an adjacent one for dislocation to form a lock-washer the length of the whole sequence could be more than one turn of helix. Besides the interaction with the first disk, parts of the RNA immediately distal to the binding site itself might still be adapted, if only secondarily to some other function, to allow as fast a growth as possible in order to quickly convert the first-formed nucleation complex to a nascent rod and stabilize it.

(ii) Conversely, the part of a nucleic acid protected from digestion by a protein which interacts with it may not necessarily include all the sequence signals required to cause the interaction in the first place. Examples are some of the sequence signals in promoter regions of *E. coli* and coliphage DNA where DNA-dependent RNA polymerase protects regions of sequence containing an axis of partial twofold rotational symmetry (within which transcription is initiated), which do not however rebind to the holoenzyme (e.g. on fd DNA, Schaller, Gray & Herrman 1975). In at least one case promoter mutations as defined genetically are well outside the region of twofold symmetry (*E. coli* lac promoter, Dickson, Abelson, Barnes & Reznikoff, 1975). A rather different example concerns the 3 ribosome binding sites of R17 RNA, which once isolated rebind to ribosomes with very different affinities to those shown when they are a part of intact RNA, the coat protein cistron ribosome binding site in particular rebinding very poorly (Steitz 1973). Since this example involves a single-stranded nucleic acid the missing RNA could be responsible for maintaining the correct tertiary structure or environment for the protected RNA, rather than contributing directly to binding. In the case of the TMV RNA fragments protected by assembly the shortest bands present in the population of protected fragments could represent molecules which, once having bound a disk and caused the conformational change to a lock-washer, have only been coated over part of the sequence involved in (say) the initial binding which would thus remain partly accessible to nuclease. Other molecules could have all this region coated, and others yet more besides. In other words, although these results do not demand a model in which sequence signals over more than one turn of virus helix (50 nucleotides) control initiation of assembly, neither do they exclude one and perhaps they may be best explained by such a model since long fragments are still preferentially made at very low protein inputs.

The third point that emerges from the RNase T_1 and pancreatic RNase fingerprints and from sequencing these end products of digestion is that there are few if any possibilities for very highly regular repeating tracts of oligonucleotides. This tentative conclusion has been confirmed by all subsequent overlapping of these end products via analysis of the products of partial digestion, which has so far yielded about three quarters of the entire sequence and most of the sequence corresponding to the first 200 nucleotides. Thus, despite the fact that this sequence interacts with a protein aggregate of exceptionally high symmetry, there is no repeating feature of an equally high order in this region of the RNA, although a weaker symmetry is not ruled out.

Since the absence of an end group or of conspicuous repeating features is unexpected, one may ask if the important part of the RNA has been lost altogether during the nuclease digestion. This may be tested by attempting to rebind the isolated fragment to the protein disk. If ^{32}P

Downloaded from rstb.royalsocietypublishing.org

labelled fragments are incubated with a disk preparation, with the protein in a large excess to drive the reaction, any nucleoprotein formed can be quantitated by passing the reaction mixture through a nitrocellulose filter and measuring the radioactivity retained on the filter. Binding of the RNA is dependent on complexing with protein since free RNA does not bind. The background due to non-specific binding can be checked using 'A-protein' instead of disks under identical conditions of incubation and filtration, and while not as low as in the complete absence of protein is still substantially lower than in the presence of disks. This method, although empirical, has been widely and successfully used for some time to study other protein–nucleic acid interactions. Its main advantage is sensitivity, and in this case binding of the fragment is linear and quantitative down to an RNA concentration (using a modal value for the chain length of 250) of 5×10^{-9} M. Lower concentrations could be tested with RNA of a higher specific activity. An obvious drawback is that one cannot test directly whether dislocation and assembly has occurred, or merely binding. However, the complex is RNase resistant. One can also attempt to measure rates of binding, although fast rates cannot conveniently be measured. Under the conditions of incubation used (5 mg/ml protein), nucleation of assembly should proceed with a half time of about 2 s (Butler 1974). Rebinding of the fragment is at least 75% complete within the minimum time necessary to do the assay, which is about 30–45 s. There are indications of further binding then occurring over a longer period to give quantitative binding by 6–10 min, however this result is preliminary and the meaning of the slow component, if any, is unclear. It could represent molecules within the population that have lost part of the information which directs binding, or a component due to non-specific binding, or a genuinely slow overall reaction. With this one reservation, such controls as we have so far performed using this method are encouraging.

(*b*) *Properties of the nucleoprotein complexes*

The nucleoprotein complexes present in the reaction mixture appear in the electron microscope as a collection of short rods of various lengths, but with a length distribution longer than that expected from the RNA fingerprint complexity (figure 5, plate 30). The complexes prepared by the standard method also show a heterogeneous and unpredictably variable pattern of sedimentation when analysed on sucrose gradients, a method that we originally hoped to use for their isolation and fractionation. Analysis of RNA extracted from both the heavy and light sides of the gradient showed that the fingerprints were identical. Therefore this apparent extreme size heterogeneity cannot be due simply to extensive asynchrony of initiation, leading to a wide range of rod sizes, but must be due to aggregation. Examination of the electron micrographs reveals a substantial number of cases in which apparently long rods are composed of two or more short ones stuck together end to end and annealed with various degrees of perfection. Since there is no evidence from analysis of the RNA chain length on gels for the presence of anything but a very small minority of the whole population of rods corresponding to chain lengths greater than 500 nucleotides, we feel that this is compelling evidence that most or all of the long rods visible in the electron microscope or as fast sedimenting material on sucrose gradients are actually end to end aggregates of short ones. We have undertaken a limited search to find conditions in which this aggregation can be eliminated, in order to fractionate the complexes at the level of nucleoprotein, but have so far been unsuccessful.

In caesium chloride buoyant density gradients the isolated nucleoprotein complexes band

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

moderately sharply with a density very slightly less than that of native TMV. This density difference can be explained by assuming that RNA tails have been partly or completely digested off both ends of the complexes, to leave a ring of protein at either end. Thus, if one of the end groups were up to 50 nucleotides away from the ends of the fragment as isolated, but lying on the surface as suggested in the previous section, it would probably have been digested away. As already noted, this begs the question of how the end group is eventually protected, nor is there any evidence that it is not more than 50 nucleotides away.

Given a modal chain length of 250 nucleotides from the nucleotide sequence data, six helix turns of protein are required to give a ring of protein at both ends, with five turns of RNA sandwiched in between. This agrees well with the presence of three disks worth of protein in the commonest class of product, and gives a predicted density corresponding to an RNA content one sixth less than native TMV. One can then calculate that protection of three quarters of the theoretical amount of RNA (§3*a*) would, excluding longer and shorter rods and frayed ends, utilize 90% of the input protein. Thus artefacts involving loss of protein from one or other end of the rods on digestion are limited to 10% of the length of the rod, or 25–30 nucleotides. Artefacts involving redistribution of protein are not so easily excluded, but again it is difficult to visualize how this might occur since digestion extensive enough to displace protein might be expected to destroy even more rapidly any uncoated RNA pieces to which this protein might transfer. Furthermore, the region of TMV RNA selected by the twin criteria of resistance to digestion by RNase T_1 (the nuclease used to prepare the protected fragment) and affinity for coat protein is presumably the fragment of the coat protein cistron already sequenced by Guilley, Jonard & Hirth (1974); see also preceding paper by Guilley *et al.*, which is not the same as the sequence selected by this procedure.

(*c*) *The protected RNA*

The ultimate aim of this study has been to sequence the protected fragment of RNA. This has been technically difficult because of the size of the fragment and the relatively low specific activity. However much progress has been made. Although two approaches designed to generate shorter rods failed – namely reducing the protein input and fractionating the nucleoprotein, one finds that if the re-extracted RNA is fractionated on polyacrylamide gels, the population separates into a number of bands – the exact number being a function of the digestion conditions used to generate the fragment (many faint sub-bands being generated by more extensive digestion, though the overall spread of the pattern is altered very little). The shortest set of bands on the gel that are present in moderate yield are in the range of 90–140 nucleotides chain length – approximately the amount of protection expected from the addition of a single disk if the RNA to protein ratio started at three turns of RNA to two turns of protein, and was then reduced by fraying. The shortest band so far detected (in very poor yield) had an electrophoretic mobility corresponding to a chain length of 55 nucleotides or slightly more than one turn of helix. When the stronger bands are eluted and fingerprinted one can clearly see that the sequences represented by the smaller bands form a part of the larger ones (figure 6), and attention has therefore been focused on this 'core'. Analysis of partial digestion products suggests that the larger bands represent covalent extensions of the smaller ones almost exclusively at one end. Only 20–30 nucleotides are added at the other, a length compatible with fraying around one turn of helix. Over 300 nucleotides can be shown to be present at the other end in the very largest band compared to the shortest ones. Furthermore, the larger bands

Downloaded from rstb.royalsocietypublishing.org

contain extensions through regions that are particularly sensitive to partial digestion with RNase T_1. Therefore the banding pattern of the RNA is probably a true reflection of the initially existing distribution of nucleoprotein, and is not caused by partial degradation by nuclease that has survived the purification. In other words, given that the original nucleoprotein distribution was not itself somehow distorted by the nuclease digestion, assembly probably nucleates in the region represented by the shortest set of bands and elongation then encapsidates the rest. The sequence signals for nucleation must therefore be present in this 'core' and possibly the adjacent regions. Sequence analysis of this region is very nearly complete, the sequence corresponding to figure 6*a*, plate 30 being overlapped in all but two places. Unfortunately one cannot deduce the polarity of assembly independently and directly from this data until the missing overlaps are found, because in their absence the order of the blocks of sequence is unknown. There are, however, various unusual features in the sequence.

First, there is a block of slightly more than one turn of helix (55 nucleotides) which is very rich in purines (71 % A+G) (figure 7). This region is immediately adjacent to another extremely A rich sequence, with a run of 16 nucleotides consisting of only A and U. This latter sequence is not present in the smallest gel bands, and so may or may not be functionally important in the assembly process. About 50 nucleotides (one helix turn) away from the major purine-rich region in the opposite direction is a third such region, containing the sequence AAAAAAG, which forms part of the 'constant' frayed end, and which is not always protected. (By 'constant' frayed end I mean the end which varies by 30 nucleotides at most.) The presence of so many tracts of purines within a space of 150 nucleotides is strikingly unusual. Since assembly with homopolymers only proceeds with the polypurines polyA and polyI (Fraenkel-Conrat & Singer 1964), this may well be of significance to assembly. The concentration of purines in this part of the sequence is not reflected in the protected region as a whole, since the base composition of the unfractionated population of protected RNA molecules is close to that of whole TMV RNA, except for a slight deficiency of C and a corresponding excess of A.

Secondly, the long purine-rich region is separated from the 'constant' frayed end by an approximately 50 nucleotide segment of much less unusual base composition, of which the

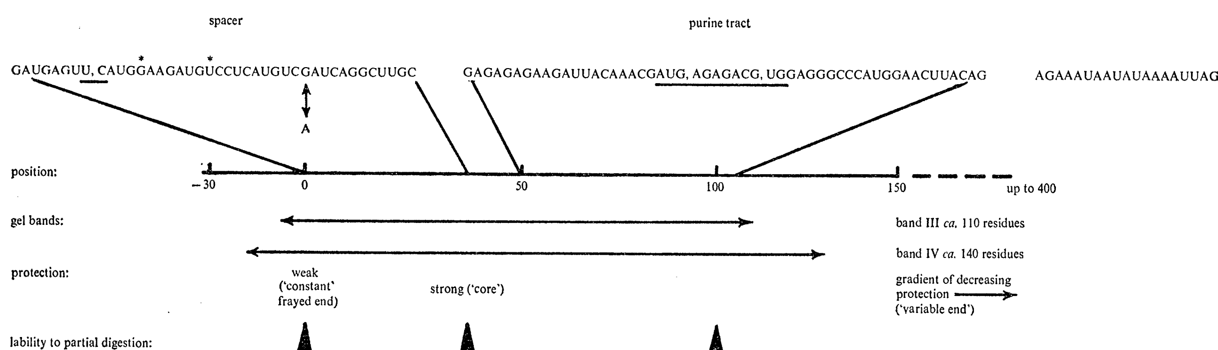


FIGURE 7. Sequences from the assembly nucleation region of TMV RNA, and the order of their protection by assembly. The sequences at the top of the figure are written conventionally, 5′ to 3′. If the direction of assembly is 5′ to 3′ they are also ordered correctly with respect to each other. If the direction is 3′ to 5′ they must be reversed in order (i.e. 5′-purine tract-spacer-constant end-3′). Positions are given with the direction of assembly taken as positive. The RNase T_1 fingerprint of band III is shown in the left panel of figure 6, that of band IV in the centre panel. A mutational difference between two stocks of TMV strain vulgare is marked. The asterisk indicates points where the dinucleotide UG may possibly be inserted, since relevant partial products were contaminated with low levels of other sequences. The oligonucleotides underlined and separated by the commas have not been ordered. The sequence given at this point is tentative.

Downloaded from rstb.royalsocietypublishing.org

major part has been sequenced (figure 7). The shortest gel band so far examined lacked part of the purine tract, but included all of this 'spacer' region. The simplest interpretation of this result is that this is the first region actually to be completely coated, whether or not it was involved in the events which led up to that coating. The broad features of the whole protected region of RNA are summarized in the figure.

If one searches the sequence for repeated features there are very few candidates. In the purine tract the commonest trinucleotide is AGA, but of these only four are at the spacing of 3n nucleotides that would be imposed by the binding of three nucleotides to each subunit. Weaker features such as a purine in every third position can be found in at least 15 triplets out of 19 in two different phases, but this is perhaps not surprising in view of the base composition. There is a tendency for features present in the first half of the tract to recur in the second half (e.g. GAGAGA... and UUACAPuA...), but not at any regular spacing. It does however appear possible to make triplet phase repeats with quite high order involving the spacer region which may be encapsidated first. Since these possible repeats may incorporate parts of the sequence outside that shown in figure 7 where the data is not yet completely firm, they will not be discussed in detail at this stage. Even here, though, the possible repeats are not straightforward triplets, but are of a fairly complex order.

The lack of strong repeating features in the sequence is perhaps not too surprising for two reasons. The first is that interactions between TMV coat protein and the RNA must of necessity be a mixture of sequence specific and sequence unspecific components. Whereas recognition of special sequence and/or secondary structure signals may be necessary for initiation of assembly, the protein must obviously bind sufficiently tightly to all other sequences in the RNA to allow coating. Over-fussiness on the part of the protein about the RNA sequences to which it would bind might unacceptably constrain those parts of the RNA that are strongly selected according to other criteria, e.g. coding sequences, replication and translation signals, etc. Thus the biological problem posed is different to that faced by (for example) a repressor or a restriction enzyme. Even in the case of repressor molecules much of the repressor present in a cell may be bound non-specifically to DNA other than the operator (Lin & Riggs 1972; Von Hippel, Revzin, Gross & Wang 1974), although since the genome of *E. coli* is 1000 times larger than TMV the competing DNA is correspondingly favoured. One wonders whether favourable binding alone can account for selection of a small part of the whole sequence for nucleation of assembly, or whether the subsequent step of dislocation is also involved in the mechanism of specificity, as suggested in the introduction. The use of a step subsequent to substrate binding to mediate catalytic specificity is known from study of enzyme mechanisms, e.g. in tRNA synthetases selecting between related amino acids (Fersht 1974).

The second reason, a corollary of the first, is that no sequence may be bound so tightly that dissociation of the protein proceeds with an inordinately long half time, otherwise the virus would not uncoat as quickly as it does (a process just as vital as assembly). Studies of the onset of increase of resistance to inactivation by u.v. light on replication in leaves (Siegel, Ginoza & Wildman 1957) show from a comparison of the lengths of eclipse after infection with either virus or RNA that uncoating takes at most of the order of 4–5 h with the strain used here. This is only one and a half orders of magnitude slower than rates measured *in vitro* for assembly (Butler 1972; Richards & Williams 1972; Butler & Finch 1973).

These considerations make questionable the desirability of the exceptionally tight binding that would be possible by tailoring the protein subunit to interact very strongly with a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

particular trinucleotide, and then supplying many repeated trinucleotides of this sort in a continuous sequence. All that is required of the nucleation sequence in the way of binding is that it must bind sufficiently strongly that subsequent steps in the mechanism proceed with a half time significantly less than the half time of dissociation of the first formed complex. Some of the preferences in base composition may reflect this requirement, as may the lack of others (note for example the scarcity of high GC regions such as are found in rRNA). However, it is by no means obvious what features of sequence or secondary structure one should look for in a mechanism designed to trigger a conformational change, although it is possible that some of the sequence regularities in the spacer are involved since this seems to be the first part of the RNA to be completely coated. Further experiments could in principle define the important region more closely, in particular the very powerful joint genetic and sequencing approach that has been used in bacterial studies (Dickson *et al.* 1975), providing, of course, that suitable mutants can be isolated. As a gesture to the future it may be worth mentioning that two different virus stocks, both TMV strain vulgare, that we have used in this study, differ in at least two places by base substitutions (both purine for purine), and one of these occurs at the position marked in the spacer (figure 7). It makes no difference we have so far detected to the properties of the virus.

4. Summary and conclusion

The experiments described above have revealed the following:

(1) The 5′ end group of TMV RNA is not a dephosphorylated adenosine residue, but a triphosphorylated guanosine, with the 5′ terminal γ-phosphate esterified with a 7-methyl guanosine blocking group. This is attached via the 5′ ribose oxygen, and carries a free 2′,3′ *cis*-hydroxyl group. The presence of the blocking group invalidates much of the indirect experimental proof for a 5′ to 3′ polar mechanism of assembly, while the remaining evidence is now equivocal.

(2) Assembly with limiting amounts of TMV protein protects a small percentage of radioactively labelled RNA from nuclease digestion. Direct sequence analysis shows that a unique region of the RNA, containing neither end group, is protected. Analysis of the isolated nucleoprotein and RNA fragments shows that even when disks are added at molar ratios with respect to RNA of close to 1:1, the most numerous class of protected RNA fragments corresponds to the addition of two or three disks. Preferential incorporation of protein into these complexes can leave more than half the input RNA totally uncoated. It is not clear how much of this preferentially protected sequence is required to direct disk binding and dislocation, though fragment rebinding experiments suggest that most or all of the fragments contain the necessary information at least to direct binding. Theoretical arguments are presented that indicate that additional features of the RNA, perhaps accounting for some of the unexpected length, may be required to direct correctly the conformational change in the protein accompanying RNA binding, rather than binding itself.

(3) Gel electrophoresis of the protected RNA shows that it consists of a population of between 50 and 450 nucleotides chain length. Fingerprinting of these bands shows that the larger ones are covalent extensions of the smaller ones, and sequence analysis shows that all but 20–30 nucleotides of this extension occurs at one end. This seems to be the direct consequence of initiation of assembly at a unique site followed by elongation predominantly in one direction. The site of the initial binding defined by the shorter gel bands has been almost

Downloaded from rstb.royalsocietypublishing.org

completely sequenced, with further information required at only two points. The polarity of assembly cannot however be directly determined without this missing information.

(4) The protected RNA does not contain very highly repetitive iterations of any one trinucleotide. It does however show a remarkable clustering of like residues in certain regions of the sequence, in particular a very purine-rich region immediately distal to the presumed disk binding site. This feature may be important to binding, dislocation, or both. Relatively complex segments of repeated sequences may be present in the region defined by the smallest gel bands and their immediately flanking regions, but exactly which of the repeats that it is possible to form is the correct one must await resolution of the remaining ambiguities in the sequence.

I would like to thank the Medical Research Council for a Research Studentship.

Note added in proof (*June* 1976). The sequence protected from nuclease digestion by limited assembly has been located between 950 and 1350 nucleotides from the 3′ end of intact TMV RNA. The predominant direction of elongation (cf. figure 7) is 3′ to 5′.

References (Zimmern)

Abraham, G., Rhodes, D. P., & Banerjee, A. K. 1975 *Nature, Lond.* **255**, 37–40.
Adams, J. M. & Cory, S. 1975 *Nature, Lond.* **255**, 28–33.
Adams, J. M., Spahr, P.-F. & Cory, S. 1972 *Biochemistry* **11**, 976–988.
Bernardi, A. & Bernardi, G. 1968 *Biochem. Biophys. Acta* **155**, 360–370.
Brownlee, G. G. 1972 In *Determination of sequences in RNA*. Amsterdam: North-Holland.
Butler, P. J. G. 1972 *J. molec. Biol.* **72**, 25–35.
Butler, P. J. G. 1974 *J. molec. Biol.* **82**, 343–353.
Butler, P. J. G. & Finch, J. T. 1973 *J. molec. Biol.* **78**, 637–649.
Butler, P. J. G. & Klug, A. 1971 *Nature New Biol.* **229**, 47–50.
Caspar, D. L. D. 1963 In *Adv. Prot. Chem.* **18**, 37–121.
Dickson, R. C., Abelson, J., Barnes, W. M. & Reznikoff, W. S. 1975 *Science, N.Y.* **187**, 27–35.
Durham, A. C. H. 1972 *J. molec. Biol.* **67**, 289–305.
Fersht, A. R. 1974 *Proc. R. Soc. Lond.* B **187**, 397–407.
Fraenkel-Conrat, H. & Fowlks, E. 1972 *Biochemistry* **11**, 1733–1736.
Fraenkel-Conrat, H. & Singer, B. 1962 *Biochemistry* **1**, 120–128.
Fraenkel-Conrat, H. & Singer, B. 1964 *Virology* **23**, 354–362.
Fraenkel-Conrat, H. & Williams, R. C. 1955 *Proc. natn. Acad. Sci., U.S.A.* **41**, 690–698.
Furuichi, Y. 1974 *Nucleic Acids Res.* **1**, 809–822.
Furuichi, Y. & Miura, K. 1975 *Nature, Lond.* **253**, 374–375.
Furuichi, Y., Morgan, M., Muthukrishnan, S. & Shatkin, A. J. 1975 *Proc. natn. Acad. Sci., U.S.A.* **72**, 362–366.
Furuichi, Y., Morgan, M., Shatkin, A. J., Jelinek, W., Salditt-Georgieff, M. & Darnell, J. R. 1975 *Proc. natn. Acad. Sci., U.S.A.* **72**, 1904–1908.
Garfin, D. E. 1972 Ph.D. Thesis, University of California, Berkeley.
Garfin, D. E. & Mandeles, S. 1975 *Virology* **64**, 388–399.
Guilley, H., Jonard, G. & Hirth, L. 1974 *Biochimie* **56**, 181–185.
Guilley, H., Stussi, C. & Hirth, L. 1971 *C. r. hebd. séanc. Acad. Sci. Paris Ser* D **272**, 1181–1184.
Hindley, J. & Staples, D. H. 1969 *Nature, Lond.* **224**, 964–967.
Kaesberg, P. 1975 In Proceedings of the INSERM-EMBO Workshop on transcription and translation of viral genomes. *INSERM* **47**, 205–210.
Lin, S-Y. & Riggs, A. D. 1972 *J. molec. Biol.* **72**, 671–690.
Miura, K., Watanabe, K. & Sugiura, M. 1974 *J. molec. Biol.* **86**, 31–48.
Ohno, T., Nozu, Y. & Okada, Y. 1971 *Virology* **44**, 510–516.
Okada, Y. & Ohno, T. 1972 *Molec. gen. Genet.* **114**, 205–213.
Perry, R. P. & Kelley, D. E. 1974 *Cell* **1**, 37–42.
Reddy, R., Ro-Choi, T. S., Henning, D. & Busch, H. 1974 *J. biol. Chem.* **249**, 6486–6494.
Richards, K. E. & Williams, R. C. 1972 *Proc. natn. Acad. Sci., U.S.A.* **69**, 1121–1124.
Richardson, C. C. 1965 *Proc. natn. Acad. Sci., U.S.A.* **54**, 158–165.

Downloaded from rstb.royalsocietypublishing.org

Rodionova, N. P., Vesenina, N. E., Kichatova, O. B. & Atabekov, J. G. 1971 *Virology* **46**, 183–191.
Schaller, H., Gray, C. & Herrman, K. 1975 *Proc. natn. Acad. Sci., U.S.A.* **72**, 737–741.
Siegel, A., Ginoza, W. & Wildman, S. G. 1957 *Virology* **3**, 554–559.
Steitz, J. A. 1969 *Nature, Lond.* **224**, 957–964.
Steitz, J. A. 1973 *Proc. natn. Acad. Sci., U.S.A.* **70**, 2605–2609.
Stussi, C., Lebeurier, G. & Hirth, L. 1969 *Virology* **38**, 16–25.
Sugiyama, T. & Fraenkel-Conrat, H. 1963 *Biochemistry* **2**, 332–334.
Suzuki, J. & Haselkorn, R. 1968 *J. molec. Biol.* **36**, 47–56.
Thouvenel, J. C., Guilley, H., Stussi, C. & Hirth, L. 1971 *FEBS Lett.* **16**, 204–206.
Tyulkina, L. G., Nazarova, G. N., Kaftanova, A. S., Ledneva, R. K., Bogdanov, A. A. & Atabekov, J. G. 1975 *Virology* **63**, 15–29.
Urushibahara, T., Furuichi, Y., Nishimura, C. & Miura, K. 1975 *FEBS Lett.* **49**, 385–389.
Von Hippel, P. M., Revzin, A., Gross, C. A. & Wang, A. C. 1974 *Proc. natn. Acad. Sci., U.S.A.* **71**, 4808–4872.
Wei, C. M., Gershowitz, A. & Moss, B. 1975 *Cell* **4**, 379–386.
Wei, C. M. & Moss, B. 1975 *Proc. natn. Acad. Sci., U.S.A.* **72**, 318–322.
Zimmern, D. 1975 *Nucleic Acids Res.* **2**, 1189–1201.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

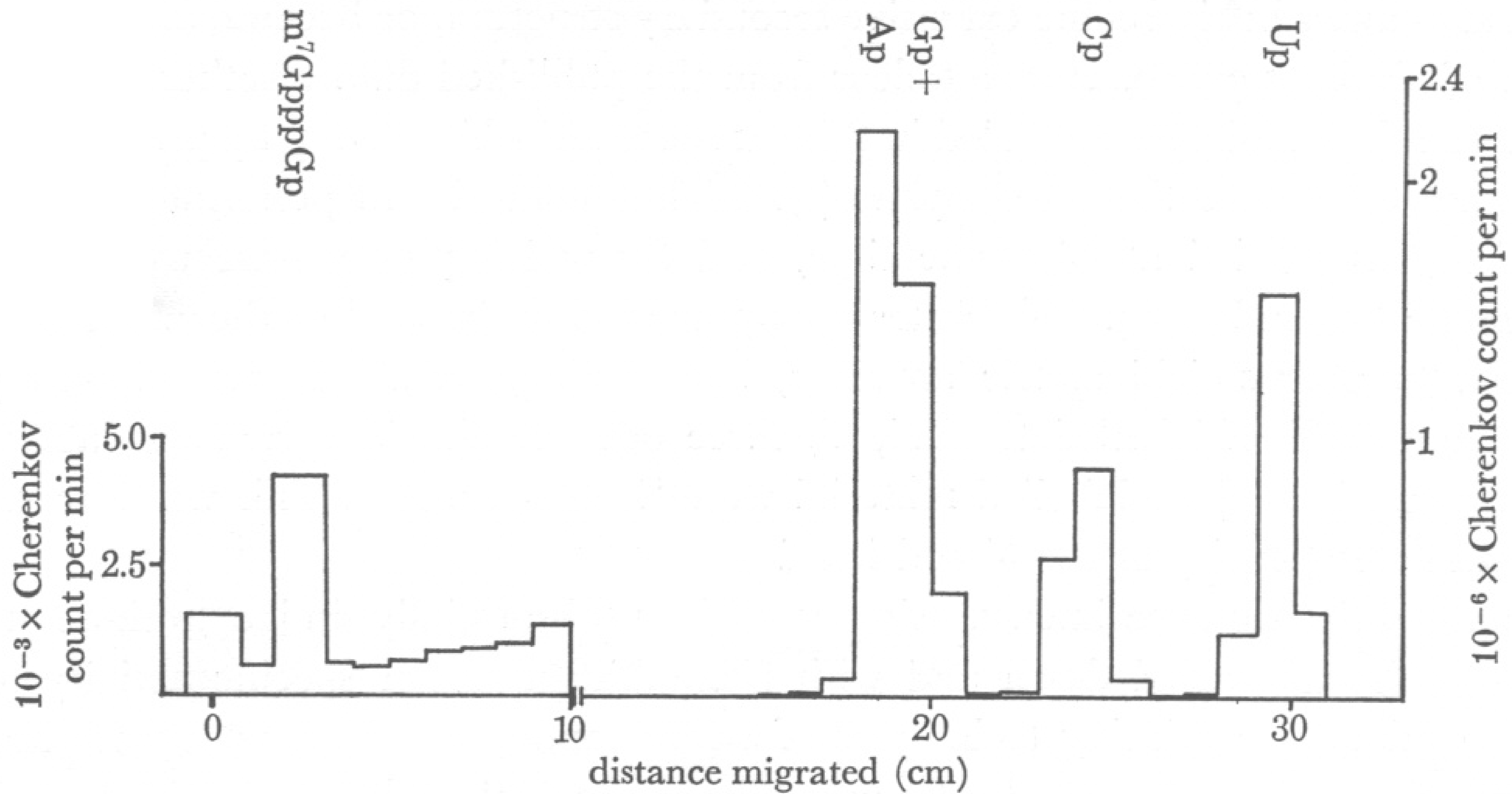
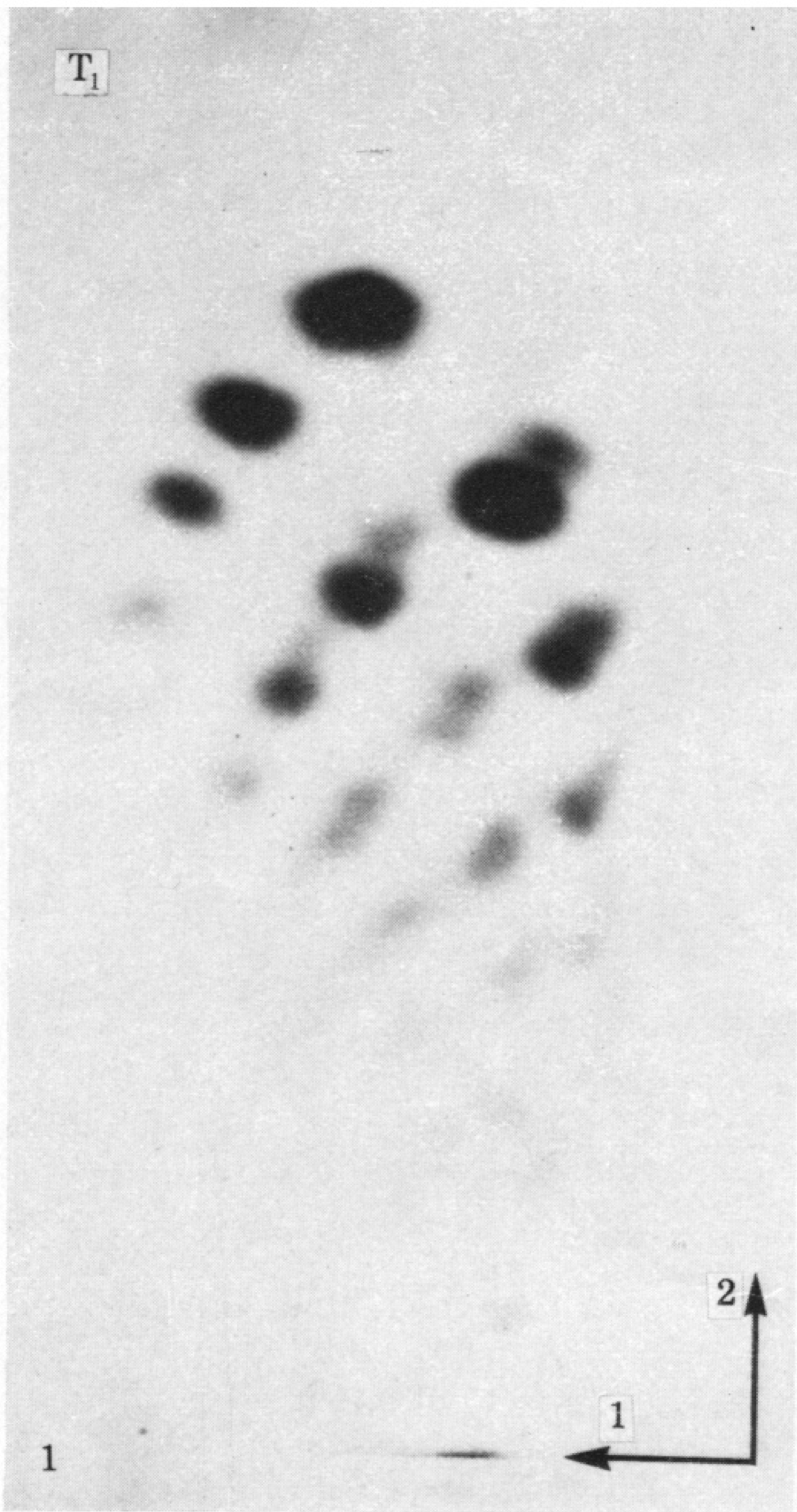
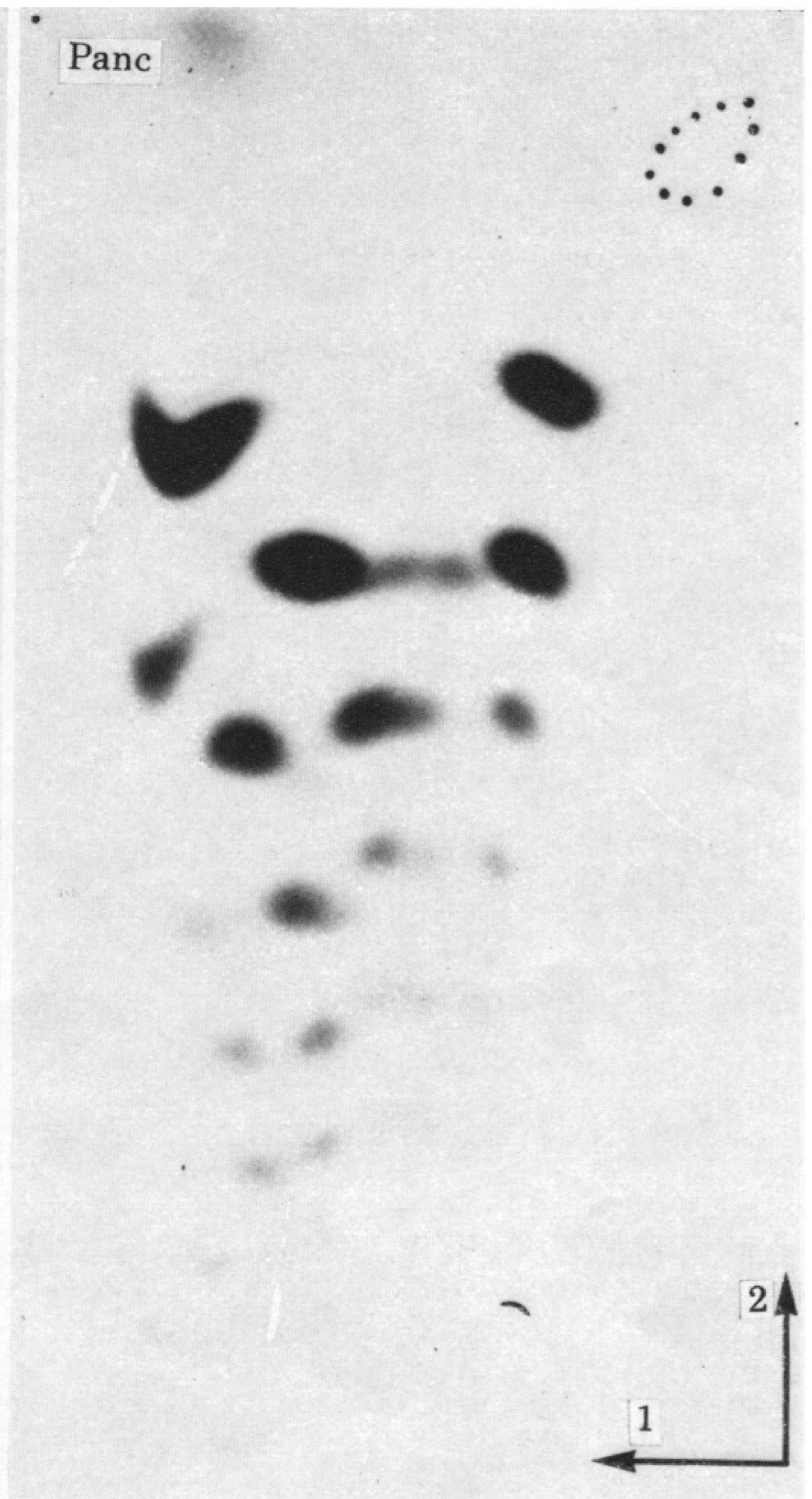


Figure 2. TMV RNA contains an oligonucleotide resistant to RNase T_2, of structure $m^7GpppGp$. An RNase T_2 digest was applied to Whatman DE 81 paper and electrophoresed at pH 3.5 (electrophoresis from left to right).

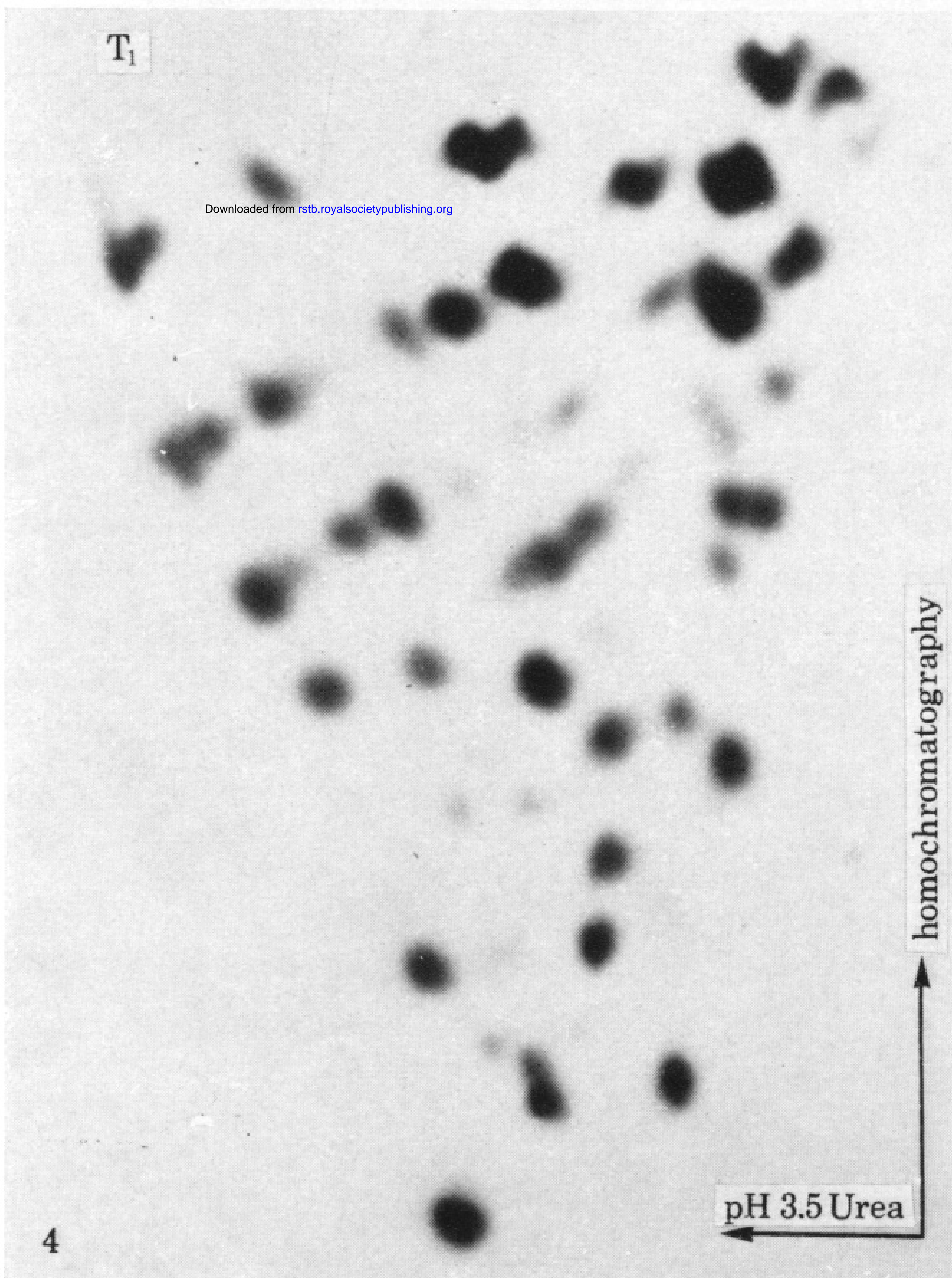
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

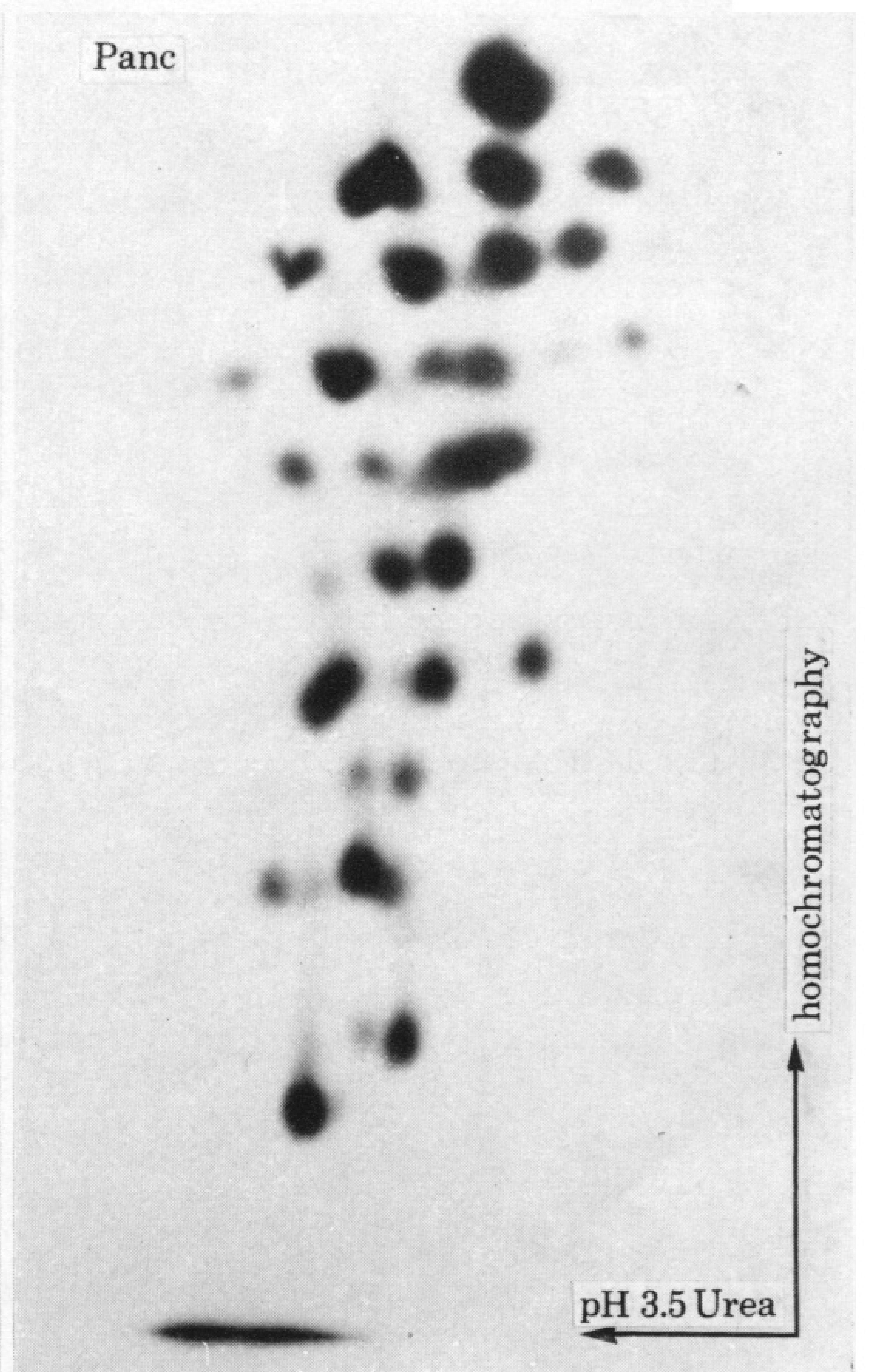
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES





Downloaded from rstb.royalsocietypublishing.org





FIGURES 1 AND 4. For description see opposite.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

FIGURE 5. Appearance in the electron microscope of material resulting from limited assembly followed by nuclease treatment. (Magnification approximately $\times 400000$.)

FIGURE 6. T_1 RNase fingerprints of bands eluted from a polyacrylamide-urea gel used to fractionate the protected RNA population. Sequences of spots are given when they first appear.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES