PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Specific Encapsidation of Fragments of TMV RNA

H. Guilley, G. Jonard, K. E. Richards and L. Hirth

Phil. Trans. R. Soc. Lond. B 1976 **276**, 181-188
doi: 10.1098/rstb.1976.0109

Phil. Trans. R. Soc. Lond. B. **276**, 181–188 (1976) [181]
Printed in Great Britain

Specific encapsidation of fragments of TMV RNA

By H. Guilley, G. Jonard, K. E. Richards and L. Hirth
Institut de Biologie Moléculaire et Cellulaire, 15 *rue Descartes*, 67000 *Strasbourg*, *France*

[Plate 28]

The *in vitro* reconstitution of tobacco mosaic virus (TMV) is initiated by the binding of a disk of TMV protein to the 'disk recognition site', a region of the RNA chain at or near the 5′-terminus for which the disk has special affinity. In order to gain insight into the recognition process, we have studied the ability of disks to encapsidate short RNA fragments produced by partial pancreatic or T_1 RNase digestion of TMV RNA. The disk is capable of dicriminating among such fragments, encapsidating only a few of the many present in the digest. The products of encapsidation are short nucleoprotein rods of the same diameter as TMV and of length proportional to that of the encapsidated RNA fragment. The particles differ from TMV, however, in one significant aspect (apart from their length): they possess rings of RNA-free protein at one or both extremities of the rod.

In the case of T_1 RNase digestion the principal encapsidated fragments were fragments T1 (105 nucleotides) and a family of smaller fragments containing elements of the same sequence. Partial digestion with pancreatic RNase generated only one major fragment (fragment P1; 150 nucleotides) with affinity for the disk. Fragment T1 has been sequenced and shown to represent a portion of the coat protein cistron. Fragment P1 has been partially sequenced but its function is not yet known. Several lines of evidence indicate that fragment T1 is *not* the disk recognition site. The portion of the TMV RNA chain from which fragment P1 is derived, on the other hand, is encapsidated early in the reconstitution process; thus fragment P1 may contain the disk recognition site.

Fragment T1 and fragment P1 both have purine-rich and cytosine-poor sequences near their termini. In addition, fragment T1, and possibly fragment P1, possess a periodicity of order three in purine residues.

It seems likely that one or both of the aforesaid properties are largely responsible for the affinity of these fragments for the disk.

The *in vitro* reconstitution of tobacco mosaic virus (TMV) from its RNA and protein components is a polar process starting at or near the 5′-terminus of the RNA (Stussi, Lebeurier & Hirth 1969; Butler & Klug 1971; Guilley, Stussi & Hirth 1971). Reconstitution is thought to be initiated by the interaction of a two-ring disk aggregate of TMV protein with the 5′-terminal region of the chain (Butler & Klug 1971). The relative spatial arrangement of the seventeen protein subunits within each ring of the disk structure is rather similar to that in the intact virus. Consequently, the subunits are well situated to interact simultaneously with as many as 51 nucleotides of the RNA chain, thus affording sufficient free energy to drive nucleation even if the individual RNA–protein interactions are weak.

Although certain nonrelated RNAs can be encapsidated by TMV protein, assembly proceeds most readily with TMV RNA. It is generally believed that the specificity of reconstitution is associated with the initiation phase of the reaction. Therefore, TMV RNA, in contrast to other RNAs, may be thought of as possessing at its 5′-end a 'recognition site' for the disk, i.e. some feature of nucleotide sequence or secondary structure for which the disk has special affinity.

In order to gain insight into the nature of the initiation reaction we have studied the ability of disks to encapsidate short RNA fragments produced by partial endonucleolytic digestion of TMV RNA. Evidently, if disks can distinguish among large RNA molecules during initiation, they should likewise be able to discriminate among small ones. In each case the basis of specificity will presumably be the same. Consequently a study of the properties of those small fragments of RNA which can become encapsidated (if such fragments exist) may be expected to shed light upon the nature of the disk recognition site. Indeed, it may prove possible to isolate the disk recognition site in this manner, provided it is not degraded by the partial enzymatic digestion.

Encapsidation of RNA fragments: properties of the nucleoprotein products of encapsidation

^{32}P TMV RNA was partially hydrolysed with T_1 RNase or pancreatic RNase and the mixture of fragments was permitted to react with a limited quantity of disks under conditions favourable for reconstitution. The encapsidated RNA fragments were separated from non-reacted RNA by sedimentation of the nucleoprotein complex. In the electron microscope the products of such reconstitution were found to consist of short rods and rings with diameters the same as that of native TMV (Jonard, Guilley & Hirth 1975). In the case of the encapsidated T_1 RNase fragments, most of the rods were about 9 nm in length, which would correspond to four turns of TMV helix. Rods of about half this length were also observed, although rather less frequently, presumably because they are rather more likely to come to rest on end upon the substrate film.

In the analytical ultracentrifuge the nucleoprotein particles formed by reconstitution with the T_1 RNase fragments sedimented as two boundaries of 26S and 34S (Jonard *et al.* 1975). The RNA extracted from the particles was also found to fall into two distinct size classes: a chain of 105 nucleotides and a family of chains of about half that length (40–60 nucleotides) (Guilley, Jonard, Richards & Hirth 1975*a*). Isopycnic gradient centrifugation revealed that the two size classes of nucleoprotein particles were the same or closely similar in RNA content (Jonard *et al.* 1975). Consequently, the chain of 105 residues must be associated with the longer nucleoprotein particle (34S) and the shorter chains with the particles of about half this size.

Tyulkina *et al.* (1975) prepared short nucleoprotein particles with physical characteristics reminiscent of those described above, using fragments of TMV RNA generated by the method of Sela (1970). The helical packing of protein subunits and the conformation of the RNA within such particles was similar to that in native virus (Tyulkina *et al.* 1975). Both Jonard *et al.* (1975) and Tyulkina *et al.* (1975) found, however, that the rodlets differ from TMV in one significant aspect (apart from their length): their RNA content is half that of native virus. Thus the 26S particle, which consists of two layers or turns of protein, has only enough RNA to fill up one of the turns; this RNA is presumably intercalated between the two layers of protein or bound in the RNA-binding groove upon one surface of the structure. Likewise, the half-complement of RNA associated with the 34S (4-turn) particle may be intercalated between the first and second and the second and third turns of the helix or between the second and third and the third and the fourth turns. In any case it is evident that half of the protein subunits in such a structure do not participate in the sort of binding to the RNA chain characteristic of native virus. Yet the nucleoprotein particles are stable under conditions where higher aggregates containing protein alone do not normally form. This finding suggests that the 'extra' protein subunits may be held in place by protein–protein interactions of an unusual sort.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

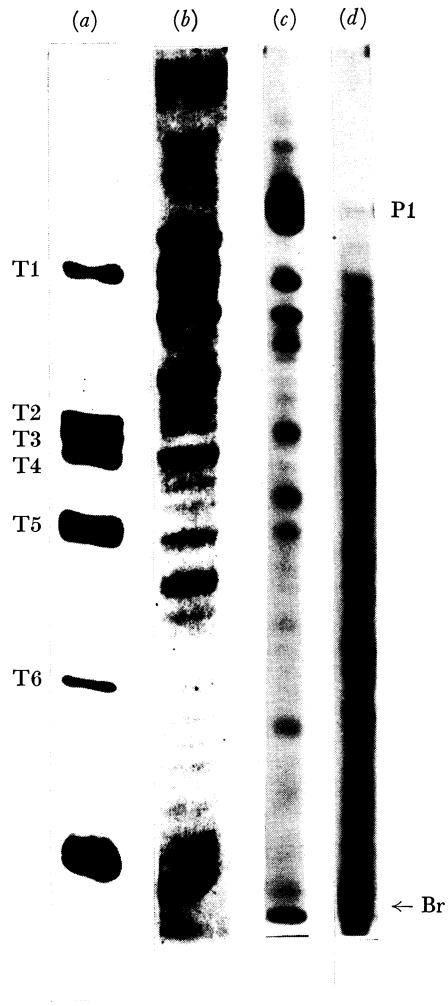


FIGURE 1. Polyacrylamide gel electrophoresis of specifically encapsidated RNA fragments generated by partial T_1 RNase digestion (*a*) and by partial pancreatic RNase digestion (*c*); the mixture of partial T_1 RNase fragments (*b*) and the mixture of partial pancreatic RNase fragments (*d*) before reconstitution. Br marks the position of the bromophenol blue marker dye.

(*Facing p.* 183)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Reaction of disks with RNA fragments produced by partial digestion with pancreatic RNase produced short nucleoprotein rods with physical properties generally similar to those obtained with the T_1 RNase fragments. In particular the ratio of protein to RNA was significantly higher than in native TMV, indicating that these particles, too, possess RNA-free rings of protein at one or both extremities.

Properties of the encapsidated RNA fragments

In the preceding section it was noted that the encapsidated RNA fragments generated by partial T_1 RNase digestion fall into two categories with respect to size: a chain of 105 residues (fragment T1) and four major fragments of about half this size (fragments T2–T5) (figure 1*a*, plate 28). Partial pancreatic RNase digestion under our conditions, on the other hand, produced only one major fragment of about 150 nucleotides (fragment P1) with affinity for the disk (figure 1*c*).

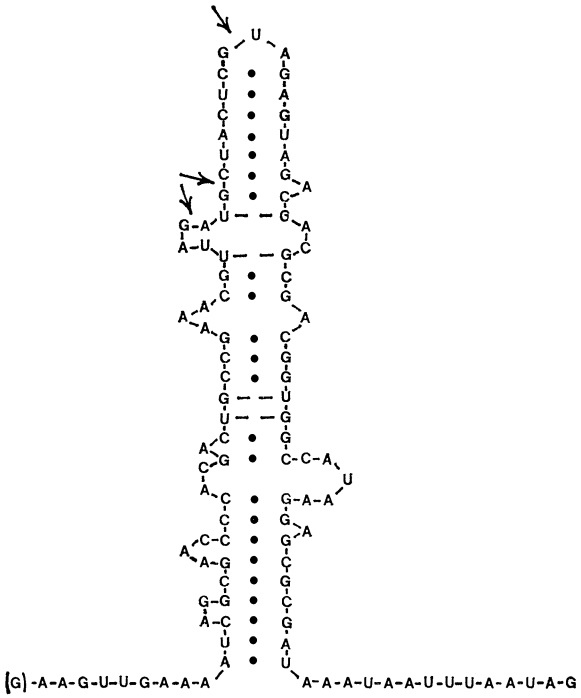


FIGURE 2. Sequence of fragment T1 folded into a possible secondary structure. The positions of the preferential scissions which produce fragments T2, T3, T4 and T5 are denoted by arrows.

Regardless of the enzyme used for digestion it is evident that the set of RNA fragments which are encapsidated is not at all representative of the spectrum of fragments found in the mixture before reconstitution (figure 1*b*, *d*). Clearly, encapsidation is selective; only certain of the many RNA fragments present in the hydrolysate can interact. Furthermore, mixed reconstitution experiments reveal that the specifically encapsidated fragments, when purified, are able to compete effectively with intact TMV RNA for disks (Jonard *et al.* 1975, and personal observations). Thus the disk-fragment reaction is not only specific but is rapid and complete.

The nucleotide sequences of the encapsidated RNA fragments have been investigated by radiochemical techniques. The sequence of fragment T1, folded into a plausible secondary structure, is shown in figure 2. It should be noted that both ends of the fragment are rich in adenosine and poor in cytidine residues. Remarkably enough, the other principal T_1 RNase

fragments which are encapsidated proved to be subproducts of fragment T1 arising from additional nucleolytic action near the middle of the chain (figure 2). Both halves of the parent fragment are represented among the encapsidated breakdown products, suggesting that the property which endows fragment T1 with affinity for the disk is not limited exclusively to one or the other extremity of the chain.

In addition to fragment T1 and its breakdown products, distinctly smaller quantities of two other T1 RNase fragments, T21 and T41, were encapsidated (Guilley, Jonard, Richards & Hirth 1975*b*). From their fingerprints it was evident that fragments T21 and T41 contain no elements of sequence in common with each other or with fragment T1. The two minor fragments were not available, however, in quantities sufficient for a complete sequence determination.

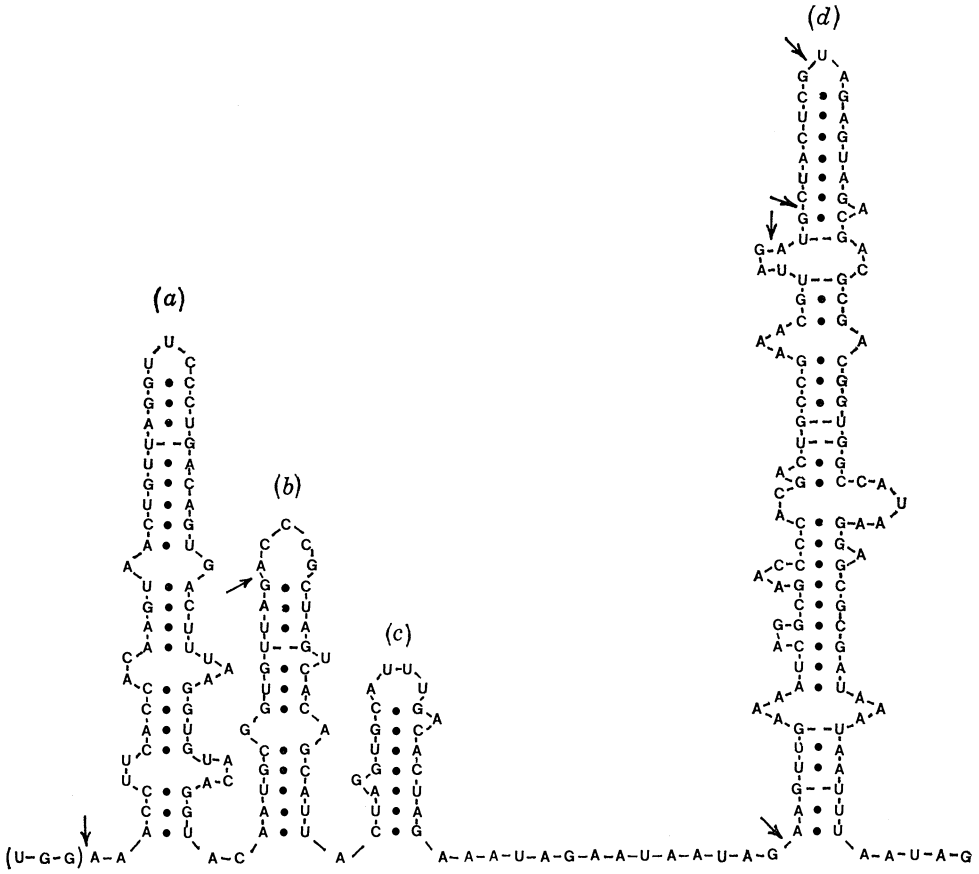


FIGURE 3. Sequence of the portion of the coat protein cistron corresponding to fragments T21–T41–T1. Arrows indicate guanylate residues susceptible to enzyme attack. The 4 hairpin loops, *a*, *b*, *c* and *d*, have stabilities of −105, −36, −23 and −133 kJ (−25.2, −8.6, −5.6 and −31.8 kcal)/mol respectively (Guilley *et al.* 1975*b*).

By analysis of its amino acid coding potential, fragment T1 was shown to originate from the TMV coat protein cistron, the portion coding for amino acids 95 to 129 of the coat protein (Richards, Guilley, Jonard & Hirth 1974). A similar analysis revealed that the two minor specifically encapsidated fragments, T21 and T41, derive from the portion of the coat protein cistron immediately adjacent to the 5′-end of fragment T1 (Guilley *et al.* 1975*b*). Together, the three fragments account for a non-interrupted stretch of 232 residues of the cistron and

contain the code for amino acids 53 to 129 of the coat protein. The order of the fragments in the sequence is T21–T41–T1 (figure 3).

In spite of its marked affinity for the disk it seems very unlikely that fragment T1 represents the disk recognition site of TMV RNA. This conclusion is based upon two lines of evidence. First, even if the coat protein cistron were situated at the 5′-terminal extremity of the RNA molecule, the part of the sequence corresponding to fragment T1 must be at least 285 residues from the 5′-end, the putative site of initiation. Indeed, we have recently shown that the fragment T1 sequence (and hence the coat protein cistron) is in fact more than 4000 nucleotides from the 5′-end (Richards *et al.* 1975). Virus were partially stripped of protein subunits starting from the 3′-end of the particle by treatment with dimethylsulphoxide and the uncoated portion of the RNA molecule was tested for the coat protein cistron by using the ability to produce fragment T1 upon partial T_1 RNase digestion as an assay. It was found that only the 3′-terminal 33% of the molecule need be uncoated in order to expose the portion of the RNA molecule from which fragment T1 is derived. Thus it would appear that the location of the T1 sequence upon the RNA molecule is incompatible with it being the disk recognition site.

A second independent argument against fragment T1 being the disk recognition site comes from partial reconstitution experiments. By definition, the site of initiation (i.e. the disk recognition site) should be the first portion of the TMV RNA chain to be encapsidated during reconstitution. Limited reconstitution between TMV RNA and protein, however, does not protect the portion of the coat protein cistron from which fragment T1 is derived from subsequent nucleolytic action (Richards *et al.* 1975). We conclude, therefore, that the portion of the RNA molecule containing the fragment T1 sequence plays no role in the initiation of reconstitution upon intact TMV RNA.

Whether or not the affinity of TMV protein for fragment T1 has a significance deeper than what would result from a chance resemblance of its sequence to that of the true disk recognition site is not apparent. It cannot be by chance, however, that the two minor specifically encapsidated fragments, T21 and T41, come from an immediately adjacent portion of the chain. We do not yet know, unfortunately, if these minor fragments do indeed have special affinity for the disk or whether they are encapsidated solely because of their proximity along the RNA molecule to fragment T1. In the latter case it would appear necessary to suppose that the three fragments have a tendency to be held together as a unit in the reconstitution medium and that encapsidation, starting with fragment T1, occasionally bridges the gaps to coat the other fragments as well.

The sequence of the long specifically encapsidated pancreatic RNase fragment, fragment P1, has not yet been completely determined. Certain aspects of its character, however, can already be discerned. In particular, fragment 1, like fragment T1, contains a high content of purines while being notably poor in cytosines near its 5′-terminus.

Fragment P1 does not contain the code for the coat protein, the only viral gene product whose primary structure is known. Consequently we have no way of establishing its function with certainty at this time. Partial reconstitution experiments with intact RNA, however, show that the P1 sequence becomes coated with protein early in the assembly process; at most, only the 5′-terminal 15% of the TMV RNA molecule need by encapsidated in order to protect the portion of the chain from which fragment P1 is derived from subsequent nucleolytic action. Thus the P1 sequence must be rather near the disk recognition site at the 5′-terminus of the RNA; it could, in fact, be the disk recognition site.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Factors influencing the recognition reaction

In the intact virus particle each protein subunit is associated with three residues of the RNA chain. Thus each subunit may be thought of as having three nucleotide binding domains. Evidently, if these binding domains have even a weak specificity toward a certain type of base residue this fact will be translated into a marked preference toward an RNA rich in that residue in the multisite recognition reaction between disk and polynucleotide. Furthermore, if the three binding domains differ in their specificities this circumstance will be manifest as a sequence specificity at the level of the disk. The sequence that is recognized will inevitably have a periodicity of order three because of the trimodal repeat in specificity of the nucleotide binding domains on the surface of the disk.

In the foregoing section we have described several small fragments of TMV RNA which are specifically encapsidated by disks. Next we shall consider what properties these fragments have in common which might account for their selection. Consideration will be limited to fragments T1 (and its breakdown products) and fragment P1 since the encapsidation of the minor fragments, T21 and T41, may be adventitious.

fragment	5′ ⟶ 3′
T1 5′	AAGUUGAAAAUCAGG...
T1 3′	...AUAAAUAAUUUAAUAG
P1 5′	AGGUUUGAGAG[AAG,AG]AUUACAAG...
T 3′OH	...CCCCGUUACCCCGGUAGGGGCCCA$_{OH}$

FIGURE 4. The purine-rich and cytidine-poor terminal regions of fragments T1 and P1; the 3′-terminus of fragment T 3′-OH, which has no affinity for the disk, is also illustrated.

In overall base composition fragment T1 and P1 do not differ markedly from total RNA. Near the 5′-extremity of fragment P1, however, is situated a stretch of 24 residues which is rich in purines (18 of 24 residues) and has a low content of cytosine (figure 4). Similar purine-rich and cytosine-poor regions are to be found at both ends of fragment T1 (figure 4). Possibly, these regions may be responsible for the high affinity of fragments P1 and T1 for the disk. Such an hypothesis would imply, of course, that one or more of the nucleotide binding sites of the protein subunit interact preferentially with purines and that consequently the recognition site must be rich in such residues. It is perhaps noteworthy that fragment T 3′-OH (Guilley, Jonard & Hirth 1975*c*) a fragment of 71 nucleotides which originates from the 3′-OH end of the intact RNA molecule and which is not recognized by the disk, is notably rich in C (figure 4). It is therefore tempting to speculate that cytidylic acid residues may exert a negative effect on the strength of the RNA–disk interaction and will consequently be underrepresented at the disk binding site.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Fragment T 3′-OH has a strong secondary structure near its 3′-terminus (Guilley *et al.* 1975*c*). The residues near the extremities of fragment T1, on the other hand, apparently do not take part in extensive base pairing (figure 2). Without doubt, the extent of base pairing and tertiary structure of a given fragment will be found to be factors of importance in determining its affinity for the disk. It is difficult to visualize, however, how such features could play other than a negative role, e.g. by 'masking' a sequence which would otherwise be favourable for interaction.

The sequence of fragment T1 possesses another characteristic which may be implicated in its recognition: a trimodal repeat in purine residues. This periodicity is illustrated in table 1 which lists the occupancy of each position of successive triplets of the fragment T1 sequence according to base composition. Eighty-three percent of the time (29 out of 35 cases) the base in the third position of each successive triplet is a purine. Although the sequence of fragment P1 is not yet completely established there are indications that it will be found to possess a similar periodicity. Evidently, if sequences of the general form PuNNPuNNPuNN...are specifically encapsidated, the TMV protein subunit must have a structure in which one of the nucleotide binding sites has a preference for purine residues.

TABLE 1. BASE COMPOSITION ACCORDING TO POSITION IN SUCCESSIVE TRIPLETS OF FRAGMENT T1

	position 1	position 2	position 3
A+G	0.42	0.63	0.83
C+U	0.58	0.37	0.17

We feel that one or perhaps both of the two aforesaid properties of fragment T1 and fragment P1 are determinent factors in regulating their affinity for the disk. This conclusion is consistent with the known preference of TMV protein, in reconstitution, for synthetic polynucleotides rich in purines, such as poly A (Fraenkel-Conrat & Singer 1964). (The affinity of poly G for TMV protein cannot be readily tested by reconstitution experiments because of the tendency for poly G to self-associate in the high ionic strength reconstitution medium.) But an explanation of the specificity of recognition which is founded solely upon an affinity for purine residues does not account for the fact that a polynucleotide consisting exclusively of purines, such as poly A, is bound less avidly by the disk than is fragment T1 (Jonard *et al.* 1975). Thus it seems likely that other features of the specifically encapsidated fragments have some as of yet unrecognized role in further strengthening the interaction.

REFERENCES (Guilley *et al.*)

Butler, P. J. G. & Klug, A. 1971 Assembly of the particle of tobacco mosaic virus from RNA and disks of protein. *Nature New Biol.* **229**, 47–50.

Fraenkel-Conrat, H. & Singer, B. 1964 Reconstitution of tobacco mosaic virus. IV. Inhibition by enzymes and other proteins, and use of polynucleotides. *Virology* **23**, 354–362.

Guilley, H., Jonard, G., Richards, K. E. & Hirth, L. 1975*a* Sequence of a specifically encapsidated RNA fragment originating from the TMV coat protein cistron. *Eur. J. Biochem.* **54**, 135–144.

Guilley, H., Jonard, G., Richards, K. E. & Hirth, L. 1975*b* Observations concerning the sequence of two additional specifically encapsidated RNA fragments originating from the TMV coat protein cistron. *Eur. J. Biochem.* **54**, 145–153.

Guilley, H., Jonard, G. & Hirth, L. 1975*c* Sequence of 71 nucleotides at the 3′-end of tobacco mosaic virus RNA. *Proc. natn. Acad. Sci., U.S.A.* **72**, 864–868.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Guilley, H., Stussi, C. & Hirth, L. 1971 Influence de la phosphodiesterase de rate de porc sur la reconstitution *in vitro* du Virus de la mosaïque du Tobac. *C. R. Acad. Sci. Paris Série D* **272**, 1181–1184.

Jonard, G., Guilley, H. & Hirth, L. 1975 Specific encapsidation of TMV RNA fragments by 25S TMV protein: Isolation and some properties of the nucleo-protein complexes formed. *Virology* **64**, 1–9.

Richards, K. E., Guilley, H., Jonard G. & Hirth, L. 1974 A specifically encapsidated fragment from the RNA of tobacco mosaic virus: sequence homology with the coat protein cistron. *F.E.B.S. Lett.* **43**, 31–32.

Richards, K. E., Morel, M. C., Nicolaïeff, A., Lebeurier, G. & Hirth, L. 1975 Location of the cistron of the tobacco mosaic virus coat protein. *Biochimie* (in the Press).

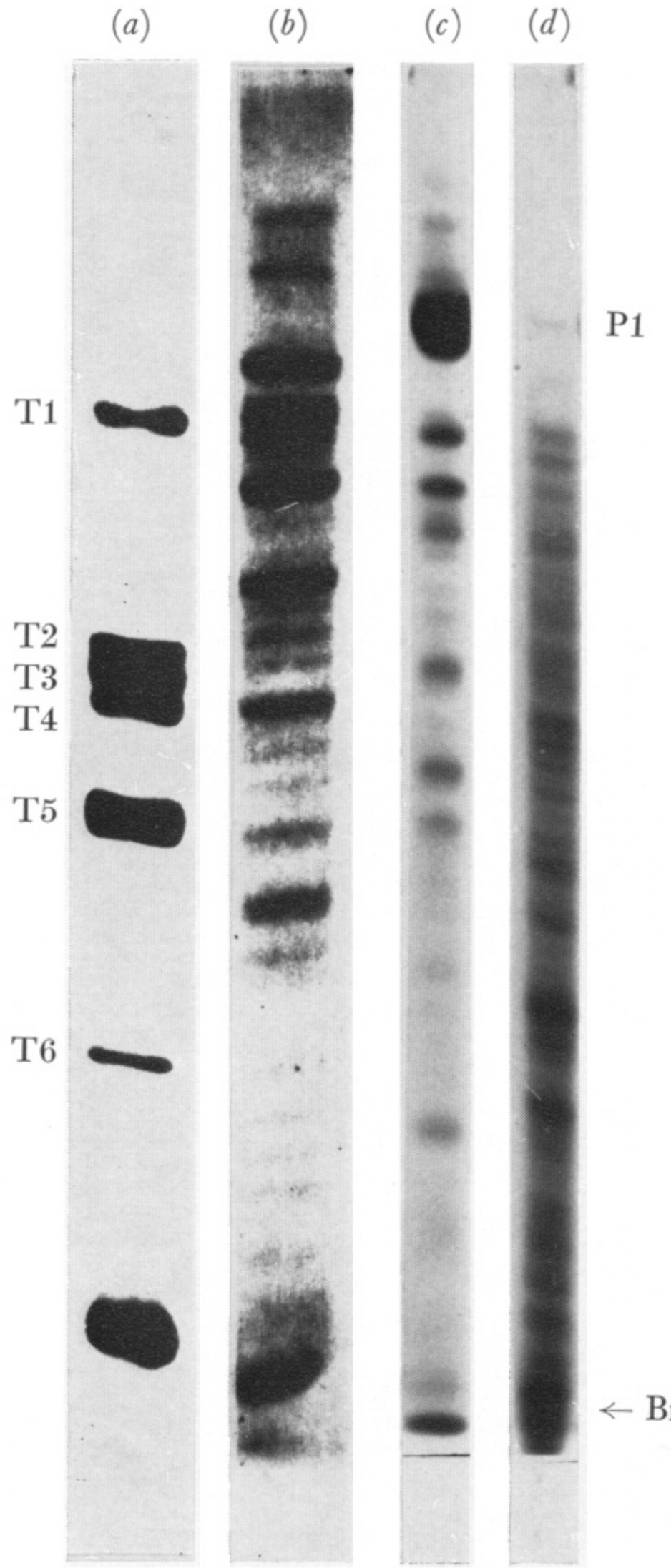
Sela, I. 1970 Controlled enzymatic cleavage of tobacco mosaic virus ribonucleic acid yielding uniform size fragments. *Arch. Biochem. Biophys.* **138**, 323–328.

Stussi, C., Lebeurier, G. & Hirth, L. 1969 Partial reconstitution of tobacco mosaic virus. *Virology* **38**, 16–25.

Thouvenel, J. C., Guilley, H., Stussi, C. & Hirth, L. 1971 Evidence for polar reconstitution of TMV. *F.E.B.S. Lett.* **16**, 204–206.

Tyulkina, L. G., Nazarova, G. N., Kaftanova, A. S., Ledneva, R. K., Bogdanov, A. A. & Atabekov, J. G. 1975 Reassembly of 20-S protein disks with 3-S RNA fragments. *Virology* **63**, 15–29.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



IGURE 1. Polyacrylamide gel electrophoresis of specifically encapsidated RNA fragments generated by partial T_1 RNase digestion (*a*) and by partial pancreatic RNase digestion (*c*); the mixture of partial T_1 RNase fragments (*b*) and the mixture of partial pancreatic RNase fragments (*d*) before reconstitution. Br marks the position of the bromophenol blue marker dye.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES