PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Self–assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed

P. J. G. Butler

Phil. Trans. R. Soc. Lond. B 1999 **354**, 537-550
doi: 10.1098/rstb.1999.0405

THE ROYAL SOCIETY

Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed

P. J. G. Butler

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The tobacco mosaic virus (TMV) particle was the first macromolecular structure to be shown to self-assemble *in vitro*, allowing detailed studies of the mechanism.

Nucleation of TMV self-assembly is by the binding of a specific stem-loop of the single-stranded viral RNA into the central hole of a two-ring sub-assembly of the coat protein, known as the 'disk'. Binding of the loop onto its specific binding site, between the two rings of the disk, leads to melting of the stem so more RNA is available to bind. The interaction of the RNA with the protein subunits in the disk cause this to dislocate into a proto-helix, rearranging the protein subunits in such a way that the axial gap between the rings at inner radii closes, entrapping the RNA.

Assembly starts at an internal site on TMV RNA, about 1 kb from its 3′-terminus, and the elongation in the two directions is different. Elongation of the nucleated rods towards the 5′-terminus occurs on a 'travelling loop' of the RNA and, predominantly, still uses the disk sub-assembly of protein subunits, consequently incorporating approximately 100 further nucleotides as each disk is added, while elongation towards the 3′-terminus uses smaller protein aggregates and does not show this 'quantized' incorporation.

Keywords: tobacco mosaic virus; self assembly; RNA–protein interactions; sub-assemblies; protein disk

1. INTRODUCTION

The tobacco mosaic virus (TMV) particle consists of a right-handed, single helix of coat protein subunits. The single-stranded RNA is intercalated between successive turns, with three nucleotides binding per protein subunit (for an early review see Caspar (1963); for a high resolution description see Namba *et al*. (1989)) and the protein completely coating the RNA. The RNA of TMV strain *vulgare* is 6395 nucleotides long (Goelet *et al*. 1982), with an inverted 7-methyl-G cap attached by a triphosphate linkage at the 5′-end (Zimmern 1975). This corresponds to approximately 2130 protein subunits and, with $16\frac{1}{3}$ subunits per helical turn, there are *ca*. 131 turns, of pitch 23 Å, making the overall particle about 300 nm long. The diameter of the helix is approximately 18 nm, with a hole of radius 2 nm up its centre, and the RNA located at a radius of 4 nm.

In some classic experiments, TMV was shown to self-assemble *in vitro* from the isolated viral RNA and protein to yield particles which both looked like native virus and had comparable infectivity (Fraenkel-Conrat & Williams 1955). This reaction was further characterized (Fraenkel-Conrat & Singer 1959) to show that it was specific for TMV RNA, in the presence of excess yeast RNA, compatible with the later observation of RNA isolated from TMV preparations contained $\leqslant 1\%$ of host plant RNA (Siegel 1971). Of the conditions tested by Fraenkel-Conrat & Singer (1959) (all with total salt at 0.1 M), the most rapid assembly from TMV RNA and the so-called 'A-protein' (a mixture of small aggregates) occurred in sodium pyrophosphate at pH 7.3 at 20 °C. With hindsight, it is worth noting that this was the highest ionic strength tested, because of the polyvalency of the pyrophosphate ions.

The specificity for TMV RNA presented a potential problem, as with only three nucleotides binding to each protein subunit, addition of single subunits during the early stages of assembly could not discriminate between RNAs. Moreover, Butler & Klug (1971) recognized that the formation of the first one or two turns of a single helix, containing an extended RNA molecule and about 16 protein subunits per turn, would be very slow. They therefore looked for a possible 'former' upon which nucleation could occur and found it in the 'disk' aggregate of the coat protein, which comprises about 80% (on a weight basis) of the 'disk preparation' produced by shifting the largely disaggregated A-protein in sodium phosphate, ionic strength 0.1 M, pH 7.0, from 4 to 20 °C and allowing the aggregation to equilibrate overnight. These 'disks' were shown to sediment at 19S and to contain about 34 subunits, which were thought to be arranged in two rings, each of 17 subunits (Durham *et al*. 1971; Durham & Klug 1971; Durham 1972*b*; Durham & Finch 1972), although there has subsequently been much controversy over the structures of *ca*. 20S protein aggregates (discussed below). It was realized that a two-layer structure would be able to assemble from single subunits much more readily than a single-start helix (Durham & Klug 1971) and it was shown that changing the conditions

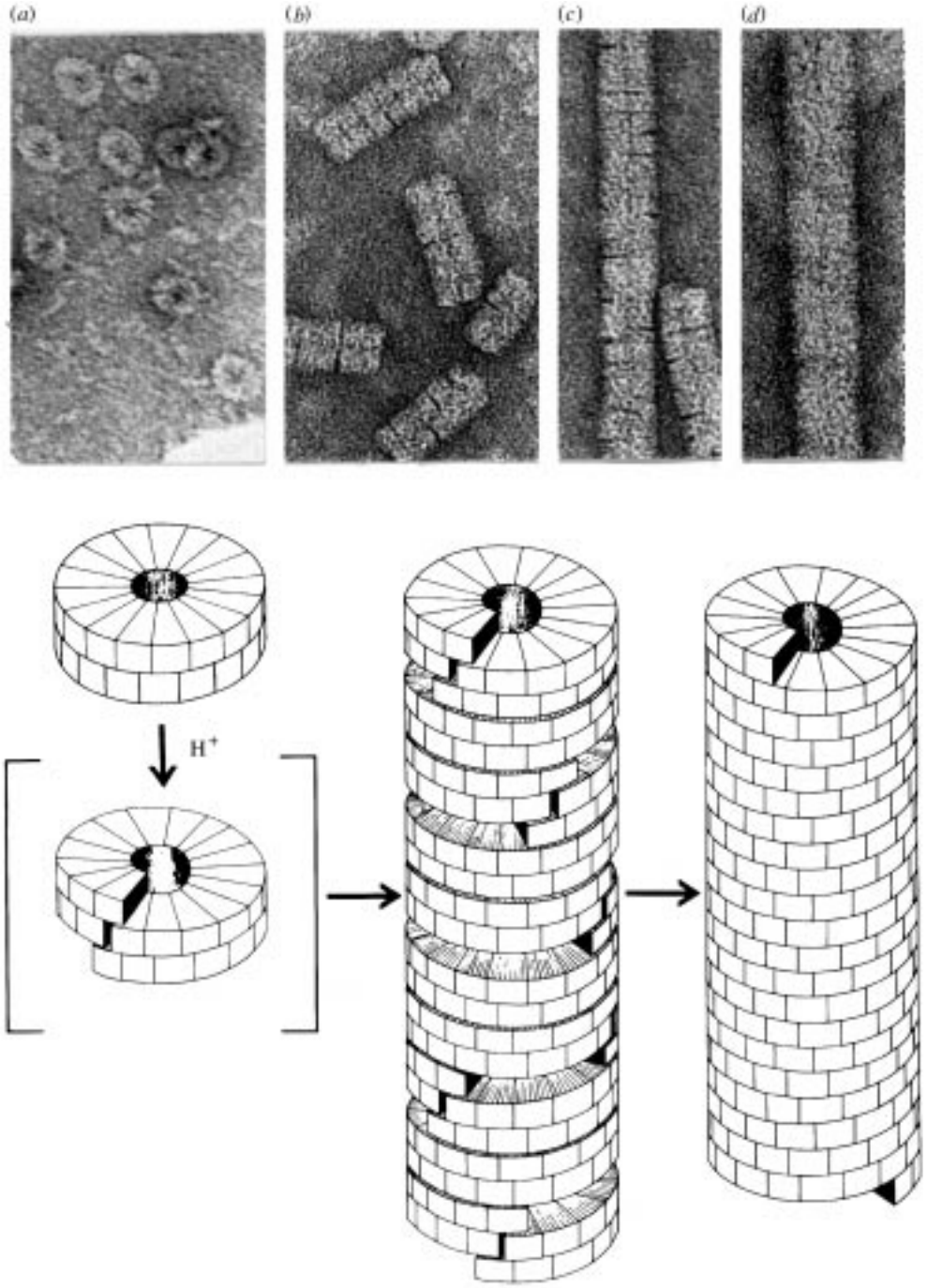

Figure 1. (*a*)–(*d*)Electron micrographs, at various times after a pH drop, of a TMV coat protein preparation. The initial 'disk preparation' was (*a*) diluted from pH 7 into pH 5 buffer (at 20 °C) and specimens prepared at (*b*) 5 s, (*c*) 15 min, (*d*) helix formed after 24 h ((*d*) is on a different scale from the other three.) A schematic interpretation of the structures is shown underneath. Electron micrographs by courtesy of Dr J. T. Finch. Adapted from Durham *et al*. (1971); Butler (1984), by permission of the Society for General Microbiology.

to those favouring the formation of the protein helix could cause these 'disks' to form long aggregates made up of discrete short helices (corresponding in size to the number of subunits in the disk), stacked in imperfect register (so-called 'nicked helices') but annealing over time into a perfect helix (figure 1; Durham *et al*. 1971). The transition from disk to imperfect helix, which takes place within a minute, does not require prior dissociation into individual subunits. Rather, the disks appear to dislocate directly into helical lockwashers.

When the effect of adding a 'disk preparation' to an assembly reaction with the A-protein was tested (Butler & Klug 1971), it was found that the rate of overall assembly was greatly increased, showing that the 20S 'disk' aggregates were playing the predicted nucleation role. However, surprisingly, the omission of added A-protein did not affect the rate of overall assembly even though this was dependent upon the concentration of the disk preparation (Butler & Klug 1971; Butler 1974*b*), i.e. the 'disk' aggregate must also have been incorporated during elongation rather than just during nucleation, even if A-protein subunits were also adding either singly or a few at a time.

These results led Butler & Klug (1971) to suggest the simple model that the 20S aggregate (which they took to

be a disk) would act during the nucleation and also subsequently during elongation. The use of such a 'sub-assembly' would have several advantages: first, the nucleating region of the RNA would be interacting with multiple protein subunits (potentially up to 17), giving the possibility of the observed specificity; second, growth from a sub-assembly allows the monitoring of subunits for their ability to assemble correctly, leading to much greater efficiency in incorporation into correctly formed final structures (Crane 1950); and finally the co-operative addition of 34 protein subunits would allow a locally unfavourable RNA sequence to be incorporated relatively readily. Other groups rapidly repeated the observation that 20S aggregates catalysed assembly of TMV, confirming their necessity for nucleation (Okada & Ohno 1972; Richards & Williams 1972), but these authors disagreed with the suggestion that they also contributed to growth. More recently, the structure of the relevant 20S aggregate has also become a matter of dispute (e.g. Correia *et al*. 1985; also reviewed by Caspar & Namba 1990). These controversies are discussed in the following sections, together with some discussion of the optimum origin of assembly on TMV RNA.

Almost all of the work has been carried out using the type (*vulgare* or U1) strain of TMV, but the group of Okada often reported results with other strains, comparing them directly with those obtained from *vulgare* without referring to the possible differences. Given the known differences in behaviour between coat proteins with only point mutations, the use of different virus strains could well lead to confusion. Similarly, one clear feature of the polymorphism of TMV coat protein is the extreme sensitivity of the aggregation to minor changes in solution conditions. For this reason it is essential to work under the same conditions if comparable results are to be obtained and experiments under different conditions can only be compared when they are carried out simultaneously, to show the effects of varying the conditions. It is particularly unsafe to compare experiments at a given ionic strength (which is expressed as a molarity) and those using buffers at the same molarity of phosphate ions—the polyvalent nature of the phosphate ion means that the ionic strength will be widely different ($\geqslant 3\times$) and strongly affect the protein aggregation.

2. STRUCTURES OF TMV COAT PROTEIN AND THE 20S AGGREGATES

(a) *X-ray structures of the protein disk and helix*

Isolated TMV coat protein at and above neutral pH had been called 'A-protein' and described as a mixture of species, predominantly around three to six subunits (Schramm & Zillig 1955). Various larger aggregates had also been described (Schramm & Zillig 1955; Anderer 1959), among which Klug & Franklin (1957) had suggested a 'disk', although what they actually described was a short proto-helical aggregate of eight hexameric A-protein aggregates. The possibility of TMV protein aggregates around 20S corresponding to a true disk-like aggregate was predicted by Caspar (1963), based upon modelling of polymorphism of aggregation for the protein. This suggested structure was similar to the asymmetric unit deduced for the crystals of TMV protein and, when a study of the aggregation over a broad range of conditions was made (Durham *et al*. 1971), the 20S aggregate was identified with the 'disk' containing two rings, each of 17 subunits (Durham *et al*. 1971; Durham & Klug 1971; Durham 1972*b*; Durham & Finch 1972). Subsequently, the crystal structure of the disk was solved to 2.8 Å resolution (Champness *et al*. 1976; Bloomer *et al*. 1978) and an atomic model built, confirming this structure in the crystal. In the crystal the polar two-ring disk forms two kinds of contact, first, stacking with another disk about a dyad axis forming a non-polar stack of disks and second, overlapping with another disk to form a 'figure-of-eight' dimer (see below). Very recently the structure has been refined further (Bhyravbhatla *et al*. 1998), but apart from locating more water molecules this has not changed the atomic model significantly. One surprise in the earlier work was the mobility of the protein at low radius and the recent model suggests that it may be less mobile when the crystals are frozen for data collection. However, NMR studies of 20S aggregates, from a TMV mutant containing a unique methionine in this 'mobile loop', showed that in solution the loop is indeed mobile in this aggregate, but not in the protein helix (Jardetzky *et al*. 1978).

The structure of TMV coat protein has also been determined both in the virus nucleoprotein helix (Stubbs *et al*. 1977; Namba & Stubbs 1986; Namba *et al*. 1989) and in the protein helix formed without RNA at low pH (Mandelkow *et al*. 1981). This latter helix is essentially identical to that in the virus and, by difference, has allowed the structure of the RNA in the virus to be determined (Stubbs & Stauffacher 1981; Namba *et al*. 1989), although inevitably with an 'average' base in each position.

(b) *Questions about the 20S aggregate in solution*

The controversy over the structure of the 20S aggregate arose from titration studies by Schuster *et al*. (1980), who interpreted their data in terms of this aggregate consisting of a mixture of 34 subunit disks and 49 subunit three-turn helices. These same workers carried out sedimentation equilibrium measurements on their 20S aggregates and, after model-fitting for the aggregation mixture, concluded that these consisted of helices with 39 ± 2 subunits (Correia *et al*. 1985). This analysis assumed that the partial specific volume was independent of aggregation state and no detail is given of how the value used was obtained. Taken together with the need to model a mixture of aggregates, one must be rather doubtful of the claimed accuracy ($\pm$*ca*. 5%) of the size determination. Moreover, it is noticeable that the S-value for the aggregate is given as 20.3S, while Durham (1972*b*) had shown that this rose abruptly from 19S to $>$20S with a small change of conditions and that the larger material could merge steadily into what were shown to be long helices. It is therefore quite probable that the aggregate(s) characterized by Schuster and his colleagues (Correia *et al*. 1985; Raghavendra *et al*. 1985, 1986, 1988) have indeed all been proto-helical, but not the same as the aggregate described as the 'disk'. Moreover, a short proto-helix would not give the distinct two-layer appearance seen in side views of the disk in the electron microscope (see below).

(c) *Stacks of disks*

Various studies have been performed on stacks of 'disks', and some of these have been interpreted as showing that the 20S aggregate in solution is not the same as the disk in the crystal. There are two distinct types of stacked disks, namely limited stacks and the so-called 'stacked disk rods' (Durham & Finch 1972). These latter are extremely stable, forming when small protein aggregates (the 'A-protein') are allowed to stand at room temperature, and have partial proteolysis of approximately 50% of the protein subunits (Carpenter 1970; Durham 1972*a*), suggesting that alternate rings will be different and contain cleaved protein subunits. While 'stacked disk rods' were useful for structural study by electron microscopy (e.g. Unwin & Klug 1974; Bloomer *et al*. 1976), the very irreversibility of their aggregation renders them a dead end. The limited stacks of disks occur at increasing ionic strength (Durham *et al*. 1971) and were considered to be related to the solution 'disk'. However, an early study (Butler & Klug 1973) showed that reassembly from these aggregates was significantly slower than that from the single 'disk', in complete agreement with the observation that they were relatively stable and converted to the helix less readily than the single 'disks' (Durham & Finch 1972). Later work has again shown a metastability of the stacks of disks (Raghavendra *et al*. 1986), so they are unlikely to be relevant to the more rapid growth of the nucleoprotein rod, although still of structural interest.

One continuing controversy concerns the polarity of these stacks. In the earliest work it was assumed that they were composed of disks, similar to the asymmetric unit of the crystal and hence polar. The stacks could therefore be either polar or non-polar, depending upon the relative orientations of adjacent disks. The stacked-disk rods had been shown to be polar (Unwin & Klug 1974) and it was generally taken that the limited stacks would be the same. However, van Regenmortel and his colleagues (Dore *et al*. 1990) reported that a monoclonal antibody, which bound to only one end of the viral helix, would bind to both ends of these limited stacks of disks, showing a non-polarity within the stack. Caspar & Namba (1990) interpreted this as showing that the disk present in these stacks must be non-polar. Subsequently stacks of disks have been examined using cryo-electron microscopy by Diaz-Avalos & Caspar (1998) and they interpreted the structure as containing a non-polar disk. Caspar further interpreted the disks stacked in his model as comprising the central, dyad-related two rings of the crystals rather than the outer rings which are the asymmetric unit.

(d) *The crystallizing aggregate*

Evidence for the probable identity of the smaller '20S' (i.e. truly 19S) solution aggregate with the disk seen in the asymmetric unit in the crystals comes from several sources. Electron micrographs of the protein during the early stages of crystallization showed many disks, but also aggregation of these disks both into 'figures-of-eight', with the edge packing of the surfaces of the disk seen in the crystals, and also two-disk stacks, some of which also had this figure-of-eight interaction (J. T. Finch, personal communication). Similar structures are observed with protein from TMV strain *dahlemense*, over a wider range of conditions (figure 2; Sperling & Klug 1975), again with an aggregating unit of about 20S and now with many side views, which correspond precisely to the crystal packing of the asymmetric units, visible in the electron microscope. Thus this figure-of-eight aggregation corresponds to the crystal packing and is only compatible with the polarity of the disk component being the same as that of the asymmetric unit in crystals.

Analysis of the rotational symmetry of both individual disks and short stacks of disks, photographed in negative stain in the electron microscope, showed that they had a dominant 17-fold symmetry and also showed a distinct handedness upon rotational averaging (Crowther & Amos 1971). This 17-fold symmetry is that expected if the disk corresponds to the crystal asymmetric unit (which is polar) and could not have occurred if the basic aggregate were a proto-helix (with $16\frac{1}{3}$ subunits per turn), as has been suggested by others. Moreover, the clear handedness of these disks and stacks, when viewed in projection, requires that not only are the disks polar themselves but also the stacking has to be polar to give an overall hand to the stack when viewed from one end.

(e) *Sensitivity of aggregation to conditions*

One complication in determining a structure for the 20S aggregate is that it is by no means certain that there is only a single aggregate present. There is certainly always a mixture of smaller A-protein present with the 20S aggregate and various authors have referred to changes in structure within this larger aggregate. One study which showed the sensitivity to conditions was a cryo-electron microscopic investigation by Butler *et al*. (1992) using the 20S:4S mixtures obtained at equilibrium at 20 °C, ionic strength 0.1 M and pH either 7.0 or 6.8. In these experiments, side views of the 20S aggregates were obtained which showed either a 'two-layer' or a 'Z-like' appearance, with the ratio changing from 10:1 at pH 7.0 to 1:10 at pH 6.8, and also 'figures-of-eight' at pH 7.0 but not pH 6.8. Computation of the images expected from either the disk, taken from the asymmetric unit of the crystal, or a 34-subunit proto-helix, by reducing the resolution of the atomic models determined by X-ray analysis, identified the 'two-layer' and 'Z-like' images with these structures. This co-existence of the disk and proto-helix and the switching of the preponderance between them with only 0.2 pH unit difference shows the danger in assuming that only a specific aggregate will be available under any chosen conditions.

(f) *Overall conclusions about the 20S aggregate state*

We therefore conclude that much of the argument about the state of aggregates around 20S may have been confounded by their being mixtures. In practice, there would usually be both polar disks, which could give figure-of-eight aggregates as occur in the crystal and also might form limited stacks, and proto-helices, which would not aggregate in the same fashion but would be able to grow by addition of limited numbers of further subunits. Either species, as well as the smaller A-protein, would therefore be available to take part in nucleoprotein assembly and any hypothesis as to which is actually involved will have to depend upon the available concentrations, where known and critical, and upon which

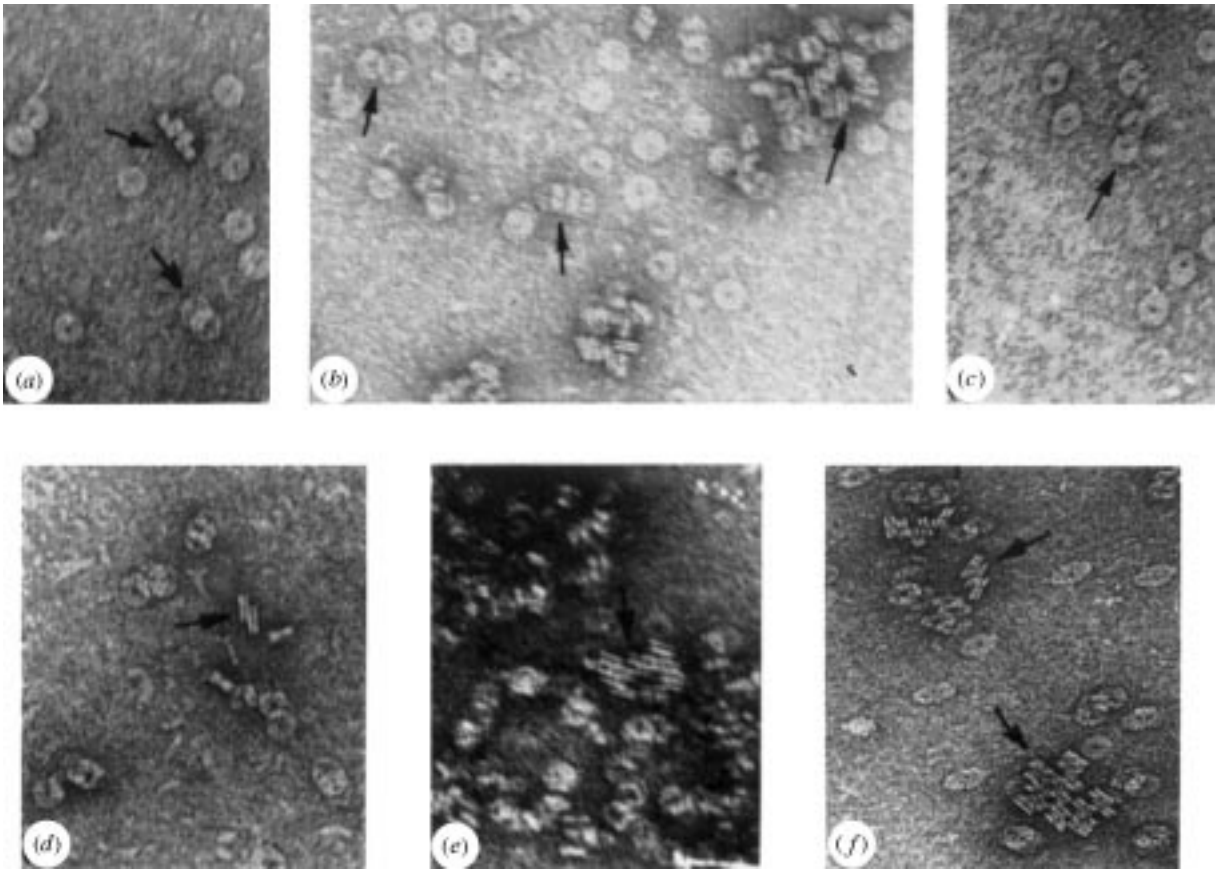

Figure 2. Electron micrographs of aggregates of protein from the *dahlemense* strain of TMV. These show the formation of higher-order aggregates, which mimic the crystal packing interactions (examples picked out with arrows), and strongly support the hypothesis that this 20S aggregate in solution (i.e. the disk) corresponds to the asymmetric unit of the crystal. Reprinted from Sperling & Klug (1975) by permission of Academic Press.

species would be the simplest to incorporate, applying usual scientific principles.

3. THE ORIGIN OF ASSEMBLY ON TMV RNA

(a) *Specificity of initiation with 20S aggregates*

The selectivity of natural encapsidation of TMV RNA has already been mentioned, with *ca*. 1% or less of the RNA being non-viral (Siegel 1971). Specificity for RNA was also observed during *in vitro* assembly, with Fraenkel-Conrat & Singer (1964) showing that the viral RNA was favoured over the other natural RNAs they tested and, of the synthetic polymers tried, only poly-A, and to a lesser extent poly-I, would assemble. With the introduction of the 20S aggregate for nucleation of assembly, Butler & Klug (1971) found a similar specificity, but now even poly-A gave only one-tenth of the rate of assembly given by TMV RNA.

(b) *Location within the RNA*

Butler & Klug (1971) observed that reassembly was prevented by prior treatment of TMV RNA with spleen phosphodiesterase, which digests exonucleolytically from the 5′-end of RNA, but not by snake venom phosphodiesterase (digesting from the 3′-end). They therefore suggested that the assembly was probably polar, with initiation at or near the 5′-terminus. This direction of elongation was supported by experiments using partially reconstituted rods (Ohno *et al*. 1971), and Guilley *et al*. (1971) confirmed the effect of spleen phosphodiesterase on reconstitution.

The first evidence to cast doubt on this interpretation came when the 5′-terminus of TMV RNA was found to have an $m^7G^{5'}ppp^{5'}Gp$ cap (Keith & Fraenkel-Conrat 1975; Zimmern 1975), making it impossible for spleen phosphodiesterase to digest it exonucleolytically, even though Zimmern had again confirmed the effect on assembly. The enzyme must therefore have contained some contaminating enzyme, which degraded the true nucleation site, wherever this was located.

Alternative approaches were used to try to identify and isolate the origin of assembly. Hirth and his colleagues partially digested TMV RNA with ribonuclease T_1 and, by binding to 20S aggregates of the protein, isolated and sequenced fragments which they assumed to be the origin (Guilley *et al*. 1974). However, subsequently they found that these coded for part of the coat protein, and concluded that this was not the origin (Guilley *et al*. 1975). Zimmern & Butler (1977) isolated a fragment of TMV RNA by allowing intact viral RNA to bind to a very limited quantity of a ‘disk preparation’ and then digesting away the majority of the RNA, before re-isolating the protected region, showing that it rebound to disks with high affinity, thus defining it as the ‘origin of assembly’. Zimmern (1977) then proceeded to sequence this fragment and, together with Wilson (Zimmern & Wilson 1976), located oligonucleotide fragments from the origin of assembly in 3′-fragments of RNA 1300 nucleotides long, but not 900, showing that the origin was about 1000

nucleotides from the 3′-end. This was subsequently confirmed when the sequence of the 3′-terminal 1000 nucleotides of TMV RNA was determined (Guilley *et al*. 1979).

The conclusion that the origin was located in the 3′-half of TMV RNA was also reached from experiments in which RNA labelled at each end, so that the individual ends could be identified, was partially reconstituted and then fragmented ultrasonically. The RNA which was encapsidated was found to contain the 3′-terminus, but not the 5′-terminus, thus locating the origin within the 3′-half of the RNA (Ohno *et al*. 1977).

The origin of assembly was sequenced both from the re-isolated fragment (Zimmern 1977) and from a longer fragment, obtained by partial ribonuclease A digestion and binding to 20S protein, which overlapped this known origin (Jonard *et al*. 1977). When the complete sequence of TMV RNA was finally determined (Goelet *et al*. 1982), the origin was shown to occur between nucleotides 5420 and 5546 in the 6395 nucleotide sequence, with the sequence AAGAAGUCG, which is suggested to form the loop in a stem-loop (figure 3) and to be the first region to bind (Zimmern 1977), located around position 5476. The high G content readily explains the absence of intact oligonucleotides from the true origin in a ribonuclease T_1 digest of TMV RNA, thus accounting for the failure to recover such fragments in the earlier work (Guilley *et al*. 1974, 1975).

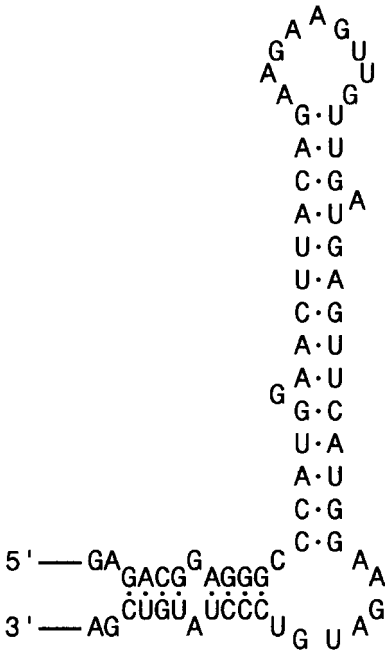

Figure 3. Nucleotide sequence and secondary structure of the origin of assembly on TMV RNA. Adapted from Zimmern (1977); reprinted from Butler (1984), by permission of the Society for General Microbiology.

(c) *Requirements for efficient nucleation*

The origin of assembly isolated by protection of RNA from nuclease digestion with limiting amounts of the 'disk preparation' amounted to more nucleotides than could be protected by a single 20S aggregate, even when protein was added at only one 'disk' per RNA molecule (Zimmern 1977; Zimmern & Butler 1977), showing that nucleation was being rapidly followed by elongation. Later Zimmern (1983) showed a longer region of rapid protection, with most of the extension in the 5′-direction from the origin. This led to a proposed secondary structure for this whole region, with the 'core' region still present as a stem-loop as described above. Cloning of the cDNA for the origin into an RNA-expression vector led both to the demonstration that *in vitro* transcribed chimaeric RNAs could be packaged into pseudo-virions (Sleat *et al*. 1986) and also to the demonstration that the presence of this region of TMV RNA alone was sufficient to allow encapsidation of contiguous foreign RNAs *in vivo* as long as TMV protein was co-expressed in the leaves (Sleat *et al*. 1988*a*,*b*). Using *in vitro* transcription of RNA from cloned DNA it was therefore possible to carry out a systematic survey of the essential requirements for efficient nucleation.

Turner & Butler (1986) found that the single stem-loop previously identified (figure 3) was both necessary and sufficient for nucleation. The sequence AAGAAGUCG, in the loop, was necessary since replacement with $(CCG)_3$ lowered the rate *ca*. 750-fold and A_9 lowered it nearly 3000-fold. Site-directed mutagenesis was then employed to see what features of the stem-loop were critical (Turner *et al*. 1988). Mutation of the loop sequence (AAGAAGUCG) to either $(UUG)_3$ or $(GUG)_3$ had little effect, but substitution of $(CCG)_3$, $(CUG)_3$ or $(UCG)_3$ was deleterious, suggesting that not only is the triplet repeat of G required, but also that C residues are unfavourable. Structural probing by chemical modification was also employed to confirm the suggested stem-loop structure. The possible role of the stem, rather than the loop, in initiation was then investigated by further mutation backed up by structural probing (Turner *et al*. 1988). This showed that substituting a perfect seven G:C base-pair tract abolished the specific initiation, although a single G:C base pair, in place of the G:U, had only a small effect.

Although the sequence of the lower half of the stem appeared relatively unimportant, the overall length of the base-paired stem was found to matter, with deletion of the bottom seven base pairs drastically lowering the rate of initiation. Deletion of the additional loop at the base of the main stem (figure 3) was shown to slow nucleation *ca*. 100-fold, but reinsertion of a different sequence fully restored it (Turner *et al*. 1988). It was speculated that the G residues in this lower loop, being in triplet phase with those in the top loop, might contribute by binding to the same position on protein subunits. However, lengthening the stem by a single base pair, which would destroy the phase, had no effect while a three base-pair addition (preserving the phase) did lower the rate *ca*. 25-fold.

The overall picture of the origin of assembly (figure 3) is that the single-stranded loop sequence (AAGAAGUCG) is essential in giving a G repeat at every third base, together with the somewhat weakly base-paired stem. The presence and distance of the lower loop contributes significantly, but its sequence is not very relevant. These features may well contribute to a low energy for the melting of the stem, readily allowing additional nucleotides to bind to protein during nucleation, once the particularly favourable loop has already bound (Butler

1984). This hypothesis for the mechanism of nucleation is discussed below.

4. THE NUCLEATION REACTION AND RESULTING STRUCTURE OF GROWING RODS

(a) *Overview*

It is generally accepted that the nucleation of TMV assembly requires a 20S aggregate, whether this be a disk or a short proto-helix. Taking this together with the location of the origin of assembly, we were led to propose the hypothesis that nucleation by insertion of the stem-loop into the central hole of a 'disk' could lead to a growing rod with both RNA tails protruding from the same end (Butler *et al*. 1976). This prediction was soon confirmed and the RNA tails were shown to protrude from that end of the growing rod which would finally encapsidate the 3′-terminus (Butler *et al*. 1977; Lebeurier *et al*. 1977). This requires the longer, 5′-tail of the RNA to be doubled back down the length of the rod, probably through the inner hole.

As well as looking at the gross structure of the rods, both of these sets of experimenters also followed the elongation. Lebeurier *et al*. (1977) followed the length of the protruding tails and concluded that, whereas the 5′-tail was encapsidated in parallel with elongation of the rod, the 3′-tail was hardly incorporated at all up to a rod length of 200 nm. Butler *et al*. (1977) compared the rate of elongation of partially assembled rods with rods prepared by partial stripping with alkali, which has been shown to expose the 5′-tail of the RNA (Perham & Wilson 1976), i.e. the longer tail, but at the opposite end compared to during elongation. Partially assembled rods were found to elongate more than an order of magnitude faster than partially stripped ones, confirming that the looped-back structure of the 5′-tail is special during assembly and that a tail simply protruding from the end of the rod is encapsidated much more slowly.

(b) *Binding of oligonucleotides to protein disks*

The earliest work reported on binding of oligonucleotides to the protein disks is that of Graham & Butler (1979), who soaked the trinucleotide AAG into protein crystals. At high concentrations, this caused the crystals to crack but at lower concentrations they were stable and X-ray diffraction data could be collected. Difference maps, from the native data, were calculated and compared to the known structure of the protein disk. The most obvious changes were shifts of various of the α-helices found in the protein (Champness *et al*. 1976; Bloomer *et al*. 1978), but in addition a positive peak, which was tentatively identified as nucleotide, was seen around the 4 nm radius and below the right radial helix, in a position which would allow the phosphate groups to interact with arginine residues 90 and 92, as observed in the virus structure (Stubbs *et al*. 1977; Namba *et al*. 1989).

Binding of oligonucleotides has also been studied by equilibrium dialysis. In the earlier work, Steckert & Schuster (1982) looked at the binding of a large number of trinucleoside diphosphates to TMV protein, finding very little binding at pH 7.0 and subsequently working at lower pH, where the protein alone would be forming helical aggregates. They found the strongest binding with AAG, with $K_d = 83\,\mu M$, and, assuming that each molecule was binding to a single protein subunit (i.e. not bridging the contact between neighbouring molecules), suggested that the G residue was favoured in the third nucleotide site on the protein. There was no evidence for this assumption nor is there still any real evidence for the preferred location of the crucial G residues, because crystallographic structures from the virus inevitably show an 'averaged' base in each position.

Turner *et al*. (1986) studied the binding of oligoribonucleotides to protein in the form of disks. They worked at pH 7.0, ionic strength 0.1 M and 20 °C, i.e. the conditions used for reassembly. They used a limited number of oligonucleotides, specifically AAG, AAGAAG and AAGAAGUUG, and, as Steckert & Schuster had done, found that AAG would not bind under these conditions. However, they did get binding of both AAGAAG and AAGAAGUUG, with K_d's of 0.9 μM and 0.7 μM, respectively, which shows stronger binding of these longer oligonucleotides at pH 7 than of the trinucleotide even at lower pH. Electron microscopy on samples prepared in parallel to the equilibrium dialysis showed that the binding of the oligonucleotides caused the formation of helices. Interestingly these showed dislocations, similar to those seen in the pH-drop experiments of Durham *et al*. (1971, 1972) (figure 1) with the 'disk preparation', and the spacing of these dislocations tended to correspond to about two turns of helix.

Turner *et al*. (1986) also looked at the potential binding of oligo-deoxyribonucleotide d(AAGAAG) and found that it bound very weakly, if at all. They estimated $K_d \geqslant 2.4\,\mu M$, compared to 0.9 μM for the ribo-oligonucleotide with this sequence, showing the importance of the 2′-hydroxyl groups in the binding. This is fully compatible with the, admittedly negative, evidence that single-stranded DNA molecules containing a cDNA sequence for the origin of assembly failed to form TMV-like helices with coat protein (Gallie *et al*. 1987).

(c) *Proposed mechanism*

The description of the origin of assembly, located about 1000 nucleotides from the 3′-end of TMV RNA, and the requirement for a 20S aggregate for assembly, led Butler *et al*. (1976) to propose the model that the stem-loop structure of the RNA would insert into the central hole of the protein aggregate, which at the time they took to be a two-ring disk, and could then interact with its binding site between the turns of the disk. The opening apart of the rings at low radius (Champness *et al*. 1976; Bloomer *et al*. 1978) would be particularly favourable in allowing access to this site. The alternative suggestion, made subsequently, that the only 20S aggregate in solution is a proto-helix faces the problem of the absence of such distortion and therefore difficulty of access to the RNA binding site.

Butler (1984) developed the model for nucleation with a disk (figure 4). He suggested that the interaction of the specific sequence in the stem, following insertion (figure 4*a*), could lead to a melting of the weakly base-paired stem, generating more single-stranded RNA which could bind to the neighbouring protein subunits until a complete turn had bound (figure 4*b*). The energy from this interaction would be sufficient to cause the necessary structural transition from disk to helix, trapping the initial

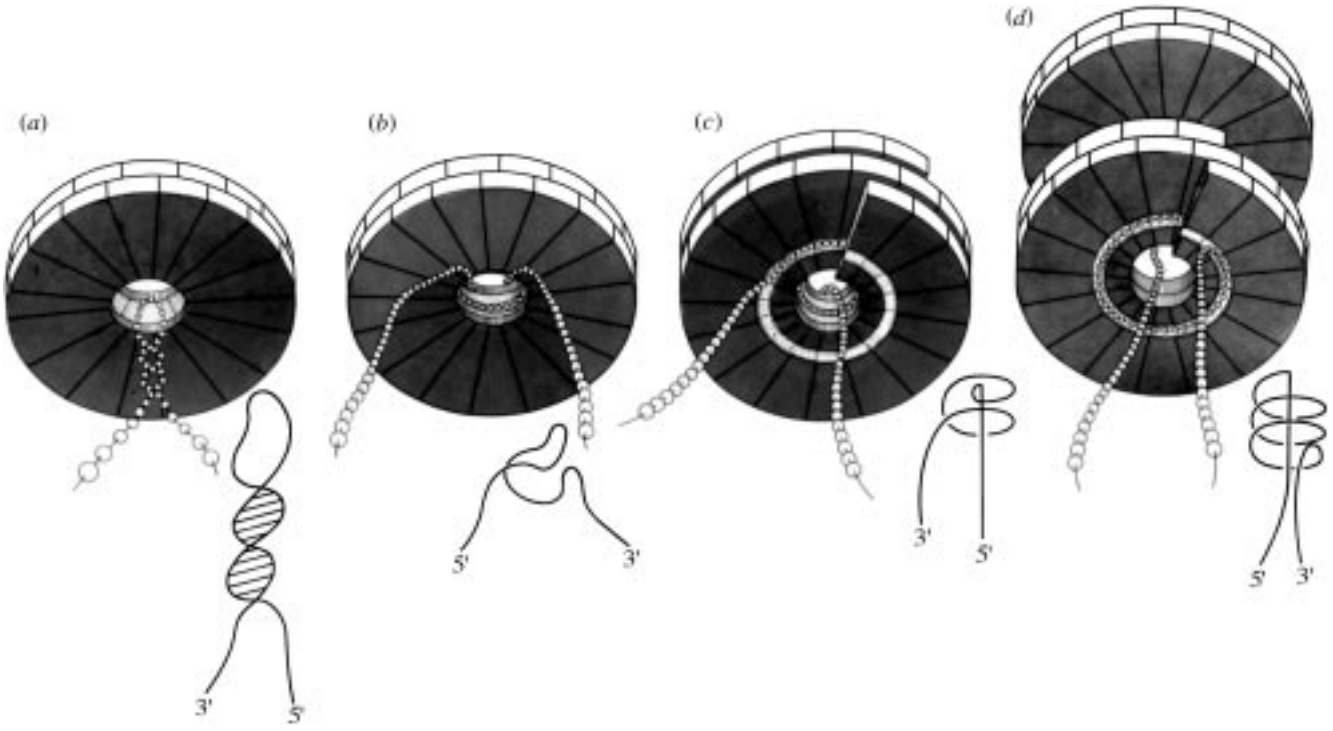

Figure 4. Picture of the possible mechanism for nucleation of assembly of TMV. The protein and RNA are drawn diagrammatically and a diagram is given for the expected configuration of the RNA backbone at each stage (see text). Reprinted from Butler (1984), by permission of the Society for General Microbiology.

RNA turn between the two layers and allowing further RNA binding onto both the top and bottom surfaces (figure 4*c*), essentially completing the nucleation reaction, with subsequent steps being elongation. The resulting structure would have the 3′-tail of the RNA coming directly from the bottom of the helix, while the 5′-tail would be doubled back down the central hole, forming a 'travelling loop' which could then interact with further 20S aggregates (figure 4*d*), by essentially the same mechanism as the nucleation (discussed below).

One point which arises is the possible co-operativity of the nucleation reaction. Most of the RNA fragments isolated from reassembly reactions with limited amounts of protein have corresponded to addition of more than one disk's-worth of protein (Zimmern 1977; Zimmern & Butler 1977) and this would be particularly likely if the nucleation were co-operative. However, when the kinetics were studied in pulse-chase experiments using radioactively labelled protein, they were best fitted by a bimolecular nucleation reaction, albeit with very rapid subsequent extension of the nucleated rodlets (Butler 1974*a*). It therefore seems likely that the nucleation is not in fact co-operative for protein aggregates but that elongation, particularly towards the 5′-end (Zimmern 1977, 1983; Turner *et al*. 1989) is very rapid and gives a similar overall picture.

5. THE ELONGATION OF NUCLEATED RODS

(a) *Bidirectional growth*

The internal location in TMV RNA of the origin of assembly (Zimmern & Wilson 1976; Guilley *et al*. 1979; Goelet *et al*. 1982) leads inevitably to nucleation within the RNA, with two tails protruding from the initial nucleoprotein helix, and the structure of these complexes has already been described with the 5′-tail looped back down the central hole and the 3′-tail protruding freely (Butler *et al*. 1977; Lebeurier *et al*. 1977). This structure means that, despite the final polar nature of the TMV rod and the natural expectation that growth would also be polar, elongation must be bidirectional and, probably, involve different mechanisms in the two directions.

It is even conceptually possible, as suggested by Okada and his colleagues (Otsuki *et al*. 1977; Fukuda *et al*. 1978), that elongation towards the 3′-end might not occur until the 5′-end was fully encapsidated, although a mechanism for preventing this is not clear. Their conclusion was based upon sequential reconstitution with protein of different strains of TMV, followed by antibody binding and electron microscopy, or on protection of labelled nucleotides at the extreme termini of the RNA. However, Lomonossoff & Butler (1979) found that oligonucleotides along the 3′-tail were protected from nuclease digestion before any of those far enough out towards the 5′-tail, showing that both tails must have been incorporated simultaneously. A similar conclusion was reached by Fairall *et al*. (1986) who showed, by hybridizing probes onto the TMV RNA and then cross-linking them so that they would block assembly at specific sites, that elongation could occur to the 3′-end when that towards the 5′-end was blocked and the tail was therefore still present. One complication in any comparison of these results and conclusions is that Okada and his colleagues worked with coat protein from different TMV strains and at a higher ionic strength than the other workers, and these facts alone could explain the differing results.

It therefore seems reasonable to accept that, for the *vulgare*-strain TMV protein at an ionic strength of 0.1 M,

elongation occurs simultaneously along both tails of the nucleated particle, although much more rapidly along the longer 5′-tail, which is the tail with the unusual configuration of a 'travelling loop', rather than a tail simply protruding from the rod. This gives a rapid encapsidation, with the protein supplied as a 'disk preparation', and full length rods and encapsidated RNA are visible after only six minutes *in vitro*. Experiments in which the elongation processes in the two directions have been studied separately enable mechanisms to be suggested for each process.

(b) *Elongation towards the 3′-terminus*

Only a limited amount of work has been done to study elongation along the 3′-tail and, because of the dominant rate in the other direction, little can be deduced from experiments where both directions of elongation can occur. In the earliest specific experiments, Lomonossoff & Butler (1980) prepared 3′-fragments of TMV RNA from rods in which the 5′-tails had been exposed by partially stripping with alkali (Perham & Wilson 1976) and removed, giving slightly over 2000 nucleotides of the 3′-end. This RNA preparation was reacted with enough of a disk preparation to coat the RNA to its 5′-ends, and further elongation from either A-protein or a disk preparation was then studied. Irrespective of whether elongation was followed as a proportion of the maximum rate achievable, or by a time-course of either the length of RNA protected from nuclease digestion or the protection of specific oligonucleotides in the 3′-tail, the rate of elongation was always faster with A-protein than with the same weight concentration of disk preparation (which contains *ca.* 20% A-protein), although saturating at much the same rate. This strongly supports the hypothesis that, towards the 3′-end, elongation is occurring by addition of subunits singly, or a few at a time, from the A-protein.

A similar conclusion came from experiments studying the lengths of RNA protected during assembly with a chimaeric RNA containing the TMV origin of assembly 5′ to a heterologous RNA sequence (Turner *et al.* 1989). The 5′-end of these chimaeric RNAs was reached within the first minute of assembly (at a rate of *ca.* 600 bases min^{-1}) and elongation then slowed markedly to *ca.* 40 bases min^{-1} as the reaction continued in the 3′-direction. Moreover, the lengths of RNA encapsidated did not show the strong *ca.* 100 base quantization seen in the 5′-direction (discussed below), but instead increased steadily, as would be compatible with addition of single subunits.

The general conclusion is therefore that elongation towards the 3′-terminus is much slower than that in the other direction, and that it probably occurs by the addition of single subunits to entrap the RNA against the end of the already formed nucleoprotein rod.

(c) *Elongation towards the 5′-terminus*

The nature of the protein aggregate involved in elongation towards the 5′-end of TMV RNA has been the subject of much controversy. When Butler & Klug (1971) first showed the involvement in assembly of the 20S aggregate, which they considered as a disk, and observed that these were used up stoichiometrically, they concluded that these disks must take part in the elongation reaction. They emphasized this rather surprising point in their discussion, to the exclusion of any comment about the possible role of A-protein even though they did consider that this would contribute under appropriate conditions (Butler 1972). Their initial omission of any discussion of a role for A-protein was taken by some other workers as implying that they thought that A-protein could not participate in the elongation (Ohno *et al.* 1972; Okada & Ohno 1972; Richards & Williams 1972), whereas these groups demonstrated that elongation could occur with A-protein as the protein source and concluded that it was the sole source.

Much of the argument has been based upon measurements of the kinetics of assembly, using a variety of techniques to determine the rate of elongation. All such measurements are complicated by the absolute necessity to initiate the reaction with at least a small amount of 20S aggregate so that it is impossible to employ pure A-protein throughout the whole experiment. Attempts have been made to get round this by pelleting partially assembled rodlets, to isolate them, and then studying the elongation of these (Richards & Williams 1973), but this still left open the question of whether the pelleting and subsequent resuspension might not have damaged the special structure of the 'travelling loop' of the 5′-tail (Butler 1974*b*).

Two different approaches have been adopted to avoid the problems inherent in such kinetics. In the first, Schuster and his colleagues added RNA to mixtures of A-protein and a disk preparation, and measured the residual free protein species by analytical centrifugation, to determine the amounts of 4S and 20S components consumed (Schuster *et al.* 1980; Shire *et al.* 1981). They found both that the protein source for $>90\%$ of the elongation could be from the 20S peak and also that a minimum of *ca.* 25% must come from this material. Although they interpreted all of the obligatory coating from 20S aggregates as nucleation, this view is debatable because such 'nucleation' would then involve over 1500 nucleotides. Nevertheless their experiments showed that much of the elongation can come from 20S aggregates, even as much as 90%, although it is not clear whether these might be breaking down to an A-protein-like state during the immediate process.

The alternative approach was by Butler and his colleagues, who looked at the lengths of RNA protected during assembly (Butler & Lomonossoff 1978; Turner *et al.* 1989). In the earlier work, Butler & Lomonossoff (1978) showed that the elongation reaction between TMV RNA and a protein 'disk preparation' could be followed by digesting away the unprotected tails of the RNA with micrococcal nuclease, at various times of assembly, and re-extracting the protected RNA. (They showed that elongation was stopped by the addition of Ca^{2+} ions needed for the nuclease digestion.) The protected RNA was sized in acrylamide–agarose gels and showed a 'quantization', with steps between successive bands in the gels averaging about 50 or 100 nucleotides. This is compatible with the addition of one or two 17-subunit rings of protein (i.e. 17×3 nucleotides per ring) at a time by direct addition of a disk aggregate (or a similar sized proto-helix).

A similar quantization of the protected RNA was seen by Schön & Mundry (1984) when they studied the

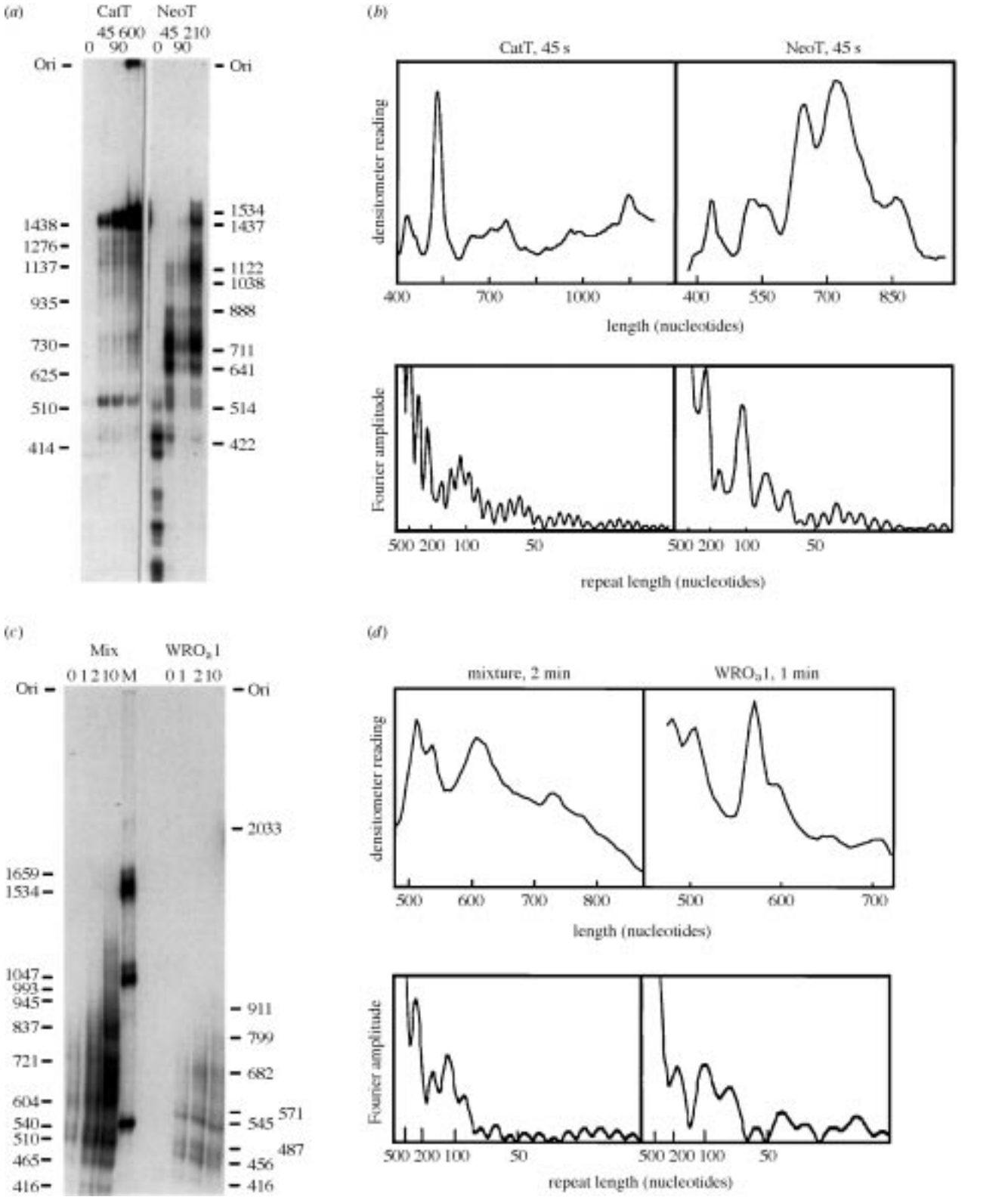

Figure 5. Analysis of protected RNA during assembly using a disk preparation and chimaeric RNAs, which have non-TMV sequences to the 5′-side of the TMV origin of assembly. (*a*) and (*b*) show experiments with hybrids containing *E. coli* RNA coding for the chloramphenicol acetyl transferase (CatT) and neomycin phosphotransferase (NeoT). (*a*) is the autoradiograph of protected RNA at different times (seconds) of assembly, with lengths marked as calculated from marker RNAs run in the same gel. (*b*) shows densitometer traces and Fourier analyses for the indicated tracks from (*a*). The entire densitometer trace shown was used for the Fourier analysis and the main peaks above the expected background correspond to repeats of 111 bases (for CatT, 45 s) and 107 bases (for NeoT, 45 s). (*c*) and (*d*) show the same as (*a*) and (*b*) (except times are now in minutes), but for experiments with either a mixture of chimaeras containing RNA coding for calf preprochymosin, chicken lysozyme and the *E. coli* neomycin phosphotransferase (Mix) or a chimaera containing RNA for the intergenic region between the 25S and 18S wheat ribosomal RNA genes (WRO$_a$1). This latter sequence contains 12 direct repeats of 135 or 136 nucleotides (Barker *et al.* 1988). The main peaks from the Fourier analysis are at 119 bases (mixed RNAs, 2 min) and 102 bases (WRO$_a$1, 1 min). Data taken from Turner *et al.* (1989).

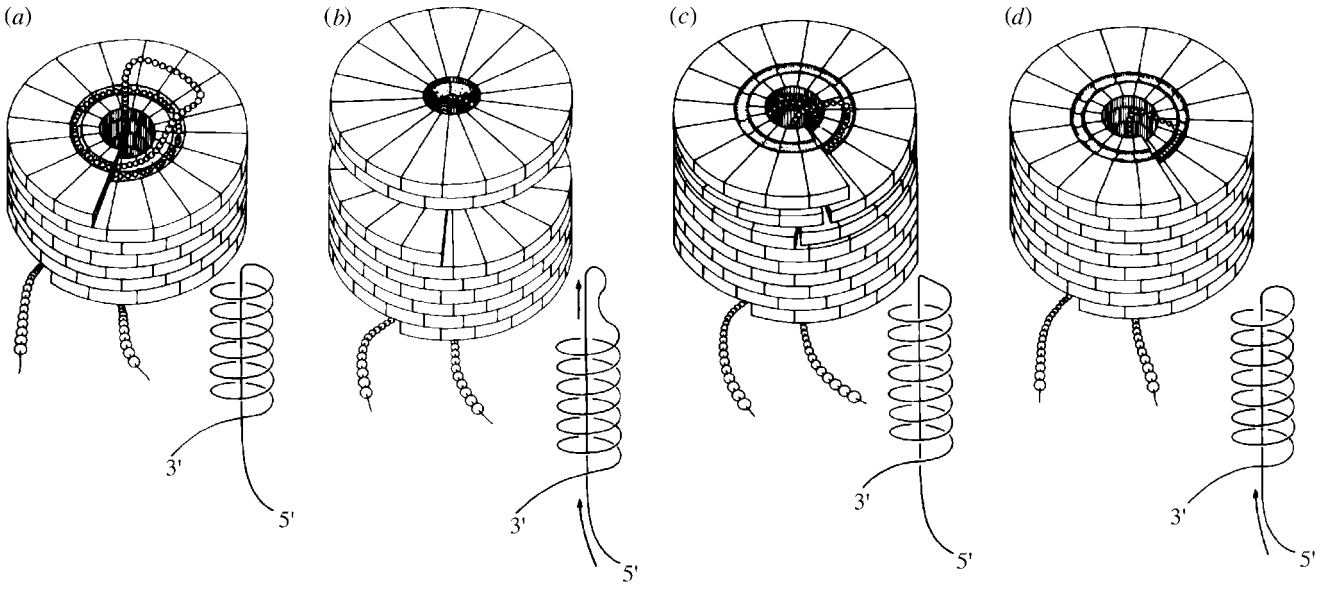

Figure 6. Picture of possible mechanism of elongation along 5′-tail from protein 'disk' (or other 20S aggregate). The protein and RNA are drawn diagrammatically and also a diagram is given for the expected configuration of the RNA and its movement up through the central hole of the rod (see text). Reprinted from Butler (1984), by permission of the Society for General Microbiology.

assembly on poly-A. When they used 20S aggregates of TMV coat protein and poly-A of about 5000 nucleotides length, they found average nucleotide increments of 188(±11) nucleotides, up to about 1000 nucleotides, and then of multimers of 102.5 nucleotides. Moreover, they showed that the 27S aggregates (the four-ring 'stack of disks') gave much slower elongation, in agreement with the finding of Butler & Klug (1973). Schön & Mundry (1984) interpreted their results as showing the addition of one or more disks at a time to the poly-A.

In contrast, Fukuda & Okada (1985) argued that the banding pattern was derived from some periodicity in the RNA sequence, rather than from the size of protein aggregate added, although they did not show any evidence for such a regular pattern in the RNA, despite the availability of the RNA sequences of several TMV strains, and it is hard to conceive what this might be in poly-A. Moreover, they again worked at much higher ionic strength than Butler and his colleagues, ignoring any possible effect of the conditions on the polymorphism of TMV protein aggregation.

In order to eliminate possible effects of any specific sequence in TMV RNA, Butler and his colleagues repeated their work and followed the elongation reaction along a variety of sequences located 5′ to the origin of assembly (Turner *et al*. 1989). These sequences came from both eukaryotic and prokaryotic sources including, in particular, the intergenic region of wheat ribosomal RNA. In order to obtain controlled protein:RNA ratios, they mixed small quantities of radioactively labelled transcript RNA with unlabelled TMV RNA as carrier. This had the added advantage that ethidium bromide staining of the gels gave an internal control for the assembly reaction occurring properly. Besides counting bands over a range of sizes, to determine the spacing of the radioactive bands, Turner *et al*. (1989) also used Fourier analysis of scans of the autoradiograms to give an unbiased measurement of the repeat frequency. In all these experiments, whether with a single RNA construct or a mixture, a dominant banding at about 100 nucleotides was observed (figure 5), with a 50 nucleotide spacing being seen on some occasions. The special significance of this length preference when using the intergenic region of wheat ribosomal RNA is that this RNA has 12 direct sequence repeats of 135 or 136 nucleotides (Barker *et al*. 1988), which would therefore dominate the elongation were this sequence dependent. While this was never the case with a 'good' disk preparation, it is noteworthy that, in one experiment where the 'disk preparation' was afterwards found to be defective (containing very little 20S aggregate), much slower elongation was seen and the repeat pattern was 142 nucleotides, i.e. close to the sequence-based repeat (Turner *et al*. 1989).

A possible model for elongation along the 5′-tail is shown in figure 6 (Butler 1984). (In these drawings the 20S aggregate has been shown as a disk.) The view shows the 'travelling loop' formed by the RNA where its 5′-end doubles back down the growing rod. This structure is probably relatively metastable and, should the elongation become too slow, could lose its ability to elongate further from disks because too much RNA had come out, giving too large a loop to participate in the rapid elongation. We suggest that the RNA binds around the top turn of the nucleoprotein helix (figure 6*a*) and then the next turn inserts into the central hole of the incoming protein aggregate, to bind to the RNA-binding site (figure 6*b*). This, together with the interaction with the end of the growing rod, causes the disk to dislocate, entrapping the RNA (figure 6*c*). Completion of this binding (figure 6*d*) will lead to the starting situation, as more RNA comes up through the central hole of the rod, to bind to the upper surface and recreate the initial situation (figure 6*a*).

6. OVERVIEW

The overall picture of the assembly of the TMV particle shows the importance of the considerable polymorphism of TMV coat protein aggregation, although

this has made elucidation of the mechanism more difficult and led to much of the controversy surrounding this. It is certain that, besides the helix formed both by coat protein alone and in the final nucleoprotein structure of the virus, the protein is needed both as a 20S aggregate, for the nucleation and fastest elongation towards the 5′-terminus of the RNA, and also as single subunits (or small aggregates, collectively known as A-protein), for the slower elongation towards the 3′-terminus.

Given the data supporting the presence of polar two-layer disks in solution under the conditions where we have seen the most rapid growth, we suggest that this 20S aggregate is involved in both nucleation and elongation. This is simpler than the alternative view that the RNA will have access between the turns of a proto-helix, when all the structural data on the helix (with or without RNA) shows the turns to be tightly bound together with no open access to the RNA binding site. The suggestion that short proto-helices will have mobile inner ends to the protein subunits is not supported by any evidence, whereas the mobile ends of the protein in the disk aggregate in both the crystal and solution have been demonstrated. The simpler hypothesis, which we suggested in 1971, still fits the available data best and is therefore preferred.

REFERENCES

Anderer, F. A. 1959 Das Molekulargewicht der Peptideinheit im Protein das Tabakmosaikvirus. *Z. Naturforsch.* **14b**, 24–28.

Barker, R. F., Harberd, N. P., Jarvis, M. G. & Flavell, R. B. 1988 Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat. *J. Mol. Biol.* **201**, 1–17.

Bhyravbhatla, B., Watowich, S. J. & Caspar, D. L. D. 1998 Refined atomic model of the four-layer aggregate of the tobacco mosaic virus coat protein at 2.4-Å resolution. *Biophys. J.* **74**, 604–615.

Bloomer, A. C., Champness, J. N. & Unwin, P. N. T. 1976 The hand of the stacked-disk aggregate of tobacco mosaic virus protein. *J. Mol. Biol.* **105**, 453–457.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. 1978 Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature* **276**, 362–368.

Butler, P. J. G. 1972 Structures and roles of the polymorphic forms of tobacco mosaic virus protein. VI. Assembly of the nucleoprotein rods of tobacco mosaic virus from the protein disks and RNA. *J. Mol. Biol.* **72**, 25–35.

Butler, P. J. G. 1974*a* Structures and roles of the polymorphic forms of tobacco mosaic virus protein. IX. Initial stages of assembly of nucleoprotein rods from viral RNA and the protein disks. *J. Mol. Biol.* **82**, 343–353.

Butler, P. J. G. 1974*b* Structures and roles of the polymorphic forms of tobacco mosaic virus protein. VIII. Elongation of nucleoprotein rods of the viral RNA and protein. *J. Mol. Biol.* **82**, 333–341.

Butler, P. J. G. 1984 The current picture of the structure and assembly of tobacco mosaic virus. *J. Gen. Virol.* **65**, 253–279.

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.

Butler, P. J. G. & Klug, A. 1973 Effect of state of polymerisation of the protein component on the assembly of tobacco mosaic virus. *Mol. Gen. Genet.* **120**, 91–93.

Butler, P. J. G. & Lomonossoff, G. P. 1978 Quantized incorporation of RNA during assembly of tobacco mosaic virus from protein disks. *J. Mol. Biol.* **126**, 877–882.

Butler, P. J. G., Bloomer, A. C., Bricogne, G., Champness, J. N., Graham, J., Guilley, H., Klug, A. & Zimmern, D. 1976 Tobacco mosaic virus assembly—specificity and the transition in protein structure during RNA packaging. In *Structure–function relationships of proteins, 3rd John Innes Symposium* (ed. R. Markham & R. W. Horne), pp. 101–110. Amsterdam: North Holland Publishing Co.

Butler, P. J. G., Finch, J. T. & Zimmern, D. 1977 Configuration of tobacco mosaic virus RNA during virus assembly. *Nature* **265**, 217–219.

Butler, P. J. G., Bloomer, A. C. & Finch, J. T. 1992 Direct visualization of the structure of the '20S' aggregate of coat protein of tobacco mosaic virus: the 'disk' is the major structure at pH 7.0 and the proto-helix at lower pH. *J. Mol. Biol.* **224**, 381–394.

Carpenter, J. M. 1970 The stacked-disk structure of tobacco mosaic virus protein. *Virology* **41**, 603–614.

Caspar, D. L. D. 1963 Assembly and stability of the tobacco mosaic virus particle. *Adv. Prot. Chem.* **18**, 37–121.

Caspar, D. L. D. & Namba, K. 1990 Switching in the self-assembly of tobacco mosaic virus. *Adv. Biophys.* **26**, 157–185.

Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. J. G. & Klug, A. 1976 The structure of the protein disk of tobacco mosaic virus to 5 Å resolution. *Nature* **259**, 20–24.

Correia, J. J., Shire, S., Yphantis, D. A. & Schuster, T. M. 1985 Sedimentation equilibrium measurements of the intermediate-size tobacco mosaic virus protein polymers. *Biochemistry* **24**, 3292–3297.

Crane, H. R. 1950 Principles & problems of biological growth. *Sci. Monthly* **70**, 376–389.

Crowther, R. A. & Amos, L. A. 1971 Harmonic analysis of electron microscope images with rotational symmetry. *J. Mol. Biol.* **60**, 123–130.

Diaz-Avalos, R. & Caspar, D. L. D. 1998 Structure of the stacked disk aggregate of tobacco mosaic virus protein. *Biophys. J.* **74**, 595–603.

Dore, I., Ruhlmann, C., Oudet, P., Cahoon, M., Caspar, D. L. D. & van Regenmortel, M. H. V. 1990 Polarity of binding of monoclonal antibodies to tobacco mosaic virus rods and stacked disks. *Virology* **176**, 25–29.

Durham, A. C. H. 1972*a* The cause of irreversible polymerisation of tobacco mosaic virus protein. *FEBS Lett.* **25**, 147–152.

Durham, A. C. H. 1972*b* Structures and roles of the polymorphic forms of tobacco mosaic virus protein. I. Sedimentation studies. *J. Mol. Biol.* **67**, 289–305.

Durham, A. C. H. & Finch, J. T. 1972 Structures and roles of the polymorphic forms of tobacco mosaic virus protein. II. Electron microscope observations of the larger polymers. *J. Mol. Biol.* **67**, 307–314.

Durham, A. C. H. & Klug, A. 1971 Polymerization of tobacco mosaic virus protein and its control. *Nature New Biol.* **229**, 42–46.

Durham, A. C. H., Finch, J. T. & Klug, A. 1971 States of aggregation of tobacco mosaic virus protein. *Nature New Biol.* **229**, 37–42.

Fairall, L., Finch, J. T., Hui, C.-F., Cantor, C. R. & Butler, P. J. G. 1986 Studies of tobacco mosaic virus reassembly with an RNA tail blocked by a hybridised and cross-linked probe. *Eur. J. Biochem.* **156**, 459–465.

Fraenkel-Conrat, H. & Singer, B. 1959 Reconstitution of tobacco mosaic virus. III. Improved methods and the use of mixed nucleic acids. *Biochim. Biophys. Acta* **33**, 359–370.

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.

Fraenkel-Conrat, H. & Williams, R. C. 1955 Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Natl Acad. Sci. USA* **41**, 690–698.

Fukuda, M. & Okada, Y. 1985 Elongation in the major direction of tobacco mosaic virus assembly. *Proc. Natl Acad. Sci. USA* **82**, 3631–3634.

Fukuda, M., Ohno, T., Okada, Y., Otsuki, Y. & Takebe, I. 1978 Kinetics of biphasic reconstitution of tobacco mosaic virus *in vitro*. *Proc. Natl Acad. Sci. USA* **75**, 1727–1730.

Gallie, D. R., Plaskitt, K. A. & Wilson, T. M. A. 1987 The effect of multiple dispersed copies of the origin-of-assembly sequence from TMV RNA on the morphology of pseudovirus particles assembled *in vitro*. *Virology* **158**, 473–476.

Goelet, P., Lomonossoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J. & Karn, J. 1982 Nucleotide sequence of tobacco mosaic virus RNA. *Proc. Natl Acad. Sci. USA* **79**, 5818–5822.

Graham, J. & Butler, P. J. G. 1979 Binding of oligonucleotides to the disk of tobacco-mosaic-virus protein. *Eur. J. Biochem.* **93**, 333–337.

Guilley, H., Stussi, C. & Hirth, L. 1971 Influence of phosphodiesterase of swine spleen on the reconstitution *in vitro* of tobacco mosaic virus. *C. R. Hebd. Scéances Acad. Sci.* Sér. D **272**, 1181–1184.

Guilley, H., Jonard, G. & Hirth, L. 1974 A TMV RNA nucleotide sequence specifically recognized by TMV protein. *Biochimie (Paris)* **56**, 181–185.

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

Guilley, H., Jonard, G., Kukla, B. & Richards, K. E. 1979 Sequence of 1000 nucleotides at the 3′ end of tobacco mosaic virus RNA. *Nucl. Acids Res.* **6**, 1287–1308.

Jardetzky, O., Akasaka, K., Vogel, D., Morris, S. & Holmes, K. C. 1978 Unusual segmental flexibility in a region of tobacco mosaic virus coat protein. *Nature* **273**, 564–566.

Jonard, G., Richards, K. E., Guilley, H. & Hirth, L. 1977 Sequence from the assembly nucleation region of TMV RNA. *Cell* **11**, 483–493.

Keith, J. & Fraenkel-Conrat, H. 1975 Tobacco mosaic virus RNA carries 5′-terminal triphosphorylated guanosine blocked by 5′-linked 7-methylguanosine. *FEBS Lett.* **57**, 31–33.

Klug, A. & Franklin, R. E. 1957 The reaggregation of the A-protein of tobacco mosaic virus. *Biochim. Biophys. Acta* **23**, 199–201.

Lebeurier, G., Nicolaieff, A. & Richards, K. E. 1977 Inside-out model for self-assembly of tobacco mosaic virus. *Proc. Natl Acad. Sci. USA* **74**, 149–153.

Lomonossoff, G. P. & Butler, P. J. G. 1979 Location & encapsidation of the coat protein cistron of tobacco mosaic virus. A bidirectional elongation of the nucleoprotein rod. *Eur. J. Biochem.* **93**, 157–164.

Lomonossoff, G. P. & Butler, P. J. G. 1980 Assembly of tobacco mosaic virus: elongation towards the 3′-hydroxyl terminus of the RNA. *FEBS Lett.* **113**, 271–274.

Mandelkow, E., Stubbs, G. & Warren, S. 1981 Structures of the helical aggregates of tobacco mosaic virus protein. *J. Mol. Biol.* **152**, 375–386.

Namba, K. & Stubbs, G. 1986 Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly. *Science* **231**, 1401–1406.

Namba, K., Pattanayek, R. & Stubbs, G. 1989 Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction. *J. Mol. Biol.* **208**, 307–325.

Ohno, T., Nozu, Y. & Okada, Y. 1971 Polar reconstitution of tobacco mosaic virus (TMV). *Virology* **44**, 510–516.

Ohno, T., Inoue, H. & Okada, Y. 1972 Assembly of rod-shaped virus *in vitro*: reconstitution with cucumber green mottle mosaic virus protein and tobacco mosaic virus RNA. *Proc. Natl Acad. Sci. USA* **69**, 3680–3683.

Ohno, T., Sumita, M. & Okada, Y. 1977 Location of the initiation site on tobacco mosaic virus RNA involved in assembly of the virus *in vitro*. *Virology* **78**, 407–414.

Okada, Y. & Ohno, T. 1972 Assembly mechanism of tobacco mosaic virus from its ribonucleic acid and protein. *Mol. Gen. Genet.* **114**, 205–213.

Otsuki, Y., Takebe, I., Ohno, T., Fukuda, M. & Okada, Y. 1977 Reconstitution of tobacco mosaic virus rods occurs bidirectionally from an internal initiation region: demonstration by electron microscopic serology. *Proc. Natl Acad. Sci. USA* **74**, 1913–1917.

Perham, R. N. & Wilson, T. M. A. 1976 The polarity of the stripping of coat protein subunits from the RNA in tobacco mosaic virus under alkaline conditions. *FEBS Lett.* **62**, 11–15.

Raghavendra, K., Adams, M. L. & Schuster, T. M. 1985 Tobacco mosaic virus protein aggregates in solution: structural comparison of 20S aggregates with those near conditions for disk crystallization. *Biochemistry* **24**, 3298–3304.

Raghavendra, K., Salunke, D. M., Caspar, D. L. D. & Schuster, T. M. 1986 Disk aggregates of tobacco mosaic virus protein in solution: electron microscopy observations. *Biochemistry* **25**, 6276–6279.

Raghavendra, K., Kelly, J. A., Khairallah, L. & Schuster, T. M. 1988 Structure and function of disk aggregates of the coat protein of tobacco mosaic virus. *Biochemistry* **27**, 7583–7588.

Richards, K. E. & Williams, R. C. 1972 Assembly of tobacco mosaic virus in vitro: effect of state of polymerization of the protein component. *Proc. Natl Acad. Sci. USA* **69**, 1121–1124.

Richards, K. E. & Williams, R. C. 1973 Assembly of tobacco mosaic virus *in vitro*. Elongation of partially assembled rods. *Biochemistry* **12**, 4574–4581.

Schön, A. & Mundry, K.-W. 1984 Coordinated two-disk nucleation, growth and properties of virus-like particles assembled from tobacco-mosaic-virus capsid protein with poly(A) or oligo(A) of different length. *Eur. J. Biochem.* **140**, 119–127.

Schramm, G. & Zillig, W. 1955 Über die Struktur des Tabakmosaikvirus. IV. Mitt.: Die Reaggregation des nucleinsäure-freien Proteins. *Z. Naturforsch.* **10b**, 493–499.

Schuster, T. M., Scheele, R. B., Adams, M. L., Shire, S. J., Steckert, J. J. & Potschka, M. 1980 Studies on the mechanism of assembly of tobacco mosaic virus. *Biophys. J.* **32**, 313–329.

Shire, S. J., Steckert, J. J. & Schuster, T. M. 1981 Mechanism of tobacco mosaic virus assembly; incorporation of 4S and 20S protein at pH 7.0 and 20 °C. *Proc. Natl Acad. Sci. USA* **78**, 256–260.

Siegel, A. 1971 Pseudovirions of tobacco mosaic virus. *Virology* **46**, 50–59.

Sleat, D. E., Turner, P. C., Finch, J. T., Butler, P. J. G. & Wilson, T. M. A. 1986 Packaging of recombinant RNA molecules into pseudovirus particles directed by the origin-of-assembly sequence from tobacco mosaic virus RNA. *Virology* **155**, 299–308.

Sleat, D. E., Gallie, D. R., Watts, J. W., Deom, C. M., Turner, P. C., Beachy, R. N. & Wilson, T. M. A. 1988*a* Selective recovery of foreign gene transcripts as virus-like particles in TMV-infected tobaccos. *Nucl. Acids Res.* **16**, 3127–3140.

Sleat, D. E., Plaskitt, K. A. & Wilson, T. M. A. 1988*b* Selective encapsidation of CAT gene transcripts in TMV-infected transgenic tobacco inhibits CAT synthesis. *Virology* **165**, 609–612.

Sperling, R. & Klug, A. 1975 States of aggregation of the *dahlamense* strain of tobacco mosaic virus protein and their relation to crystal formation. *J. Mol. Biol.* **96**, 425–430.

Steckert, J. J. & Schuster, T. M. 1982 Sequence specificity of trinucleoside diphosphate binding to polymerized tobacco mosaic virus protein. *Nature* **299**, 32–36.

Stubbs, G. & Stauffacher, C. 1981 Structure of the RNA in tobacco mosaic virus. *J. Mol. Biol.* **152**, 387–396.

Stubbs, G., Warren, S. & Holmes, K. 1977 Structure of RNA and RNA binding site in tobacco mosaic virus from 4-Å map calculated from X-ray fibre diagrams. *Nature* **267**, 216–221.

Turner, D. R. & Butler, P. J. G. 1986 Essential features of the assembly origin of tobacco mosaic virus RNA as studied by directed mutagenesis. *Nucl. Acids Res.* **14**, 9229–9242.

Turner, D. R., Mondragon, A., Fairall, L., Bloomer, A. C., Finch, J. T., Van Boom, J. H. & Butler, P. J. G. 1986 Oligonucleotide binding to the coat protein disk of tobacco mosaic virus. Possible steps in the mechanism of assembly. *Eur. J. Biochem.* **157**, 269–274.

Turner, D. R., Joyce, L. E. & Butler, P. J. G. 1988 The tobacco mosaic virus assembly origin RNA. Functional characteristics defined by directed mutagenesis. *J. Mol. Biol.* **203**, 531–547.

Turner, D. R., McGuigan, C. J. & Butler, P. J. G. 1989 Assembly of hybrid RNAs with tobacco mosaic virus coat protein. Evidence for incorporation of disks in 5′-elongation along the major RNA tail. *J. Mol. Biol.* **209**, 407–422.

Unwin, P. N. T. & Klug, A. 1974 Electron microscopy of the stacked disk aggregate of tobacco mosaic virus protein. *J. Mol. Biol.* **87**, 641–656.

Zimmern, D. 1975 The 5′ end group of tobacco mosaic virus is $m^7G^{5'}ppp^{5'}Gp$. *Nucl. Acids Res.* **2**, 1189–1201.

Zimmern, D. 1977 Nucleotide sequence at the origin for assembly on tobacco mosaic virus RNA. *Cell* **11**, 463–482.

Zimmern, D. 1983 An extended secondary structure model for the TMV assembly origin, and its correlation with protection studies and an assembly defective mutant. *EMBO J.* **2**, 1901–1907.

Zimmern, D. & Butler, P. J. G. 1977 The isolation of tobacco mosaic virus RNA fragments containing the origin for viral assembly. *Cell* **11**, 455–462.

Zimmern, D. & Wilson, T. M. A. 1976 Location of the origin for viral reassembly on tobacco mosaic virus RNA and its relation to stable fragment. *FEBS Lett.* **71**, 294–298.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES