Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The tobacco mosaic virus particle: structure and assembly

A. Klug

Phil. Trans. R. Soc. Lond. B 1999 **354**, 531-535
doi: 10.1098/rstb.1999.0404

References Article cited in:
http://rstb.royalsocietypublishing.org/content/354/1383/531#related-urls

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

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

This journal is © 1999 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

THE ROYAL SOCIETY

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The tobacco mosaic virus particle: structure and assembly

A. Klug

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

A short account is given of the physical and chemical studies that have led to an understanding of the structure of the tobacco mosaic virus particle and how it is assembled from its constituent coat protein and RNA. The assembly is a much more complex process than might have been expected from the simplicity of the helical design of the particle. The protein forms an obligatory intermediate (a cylindrical disk composed of two layers of protein units), which recognizes a specific RNA hairpin sequence. This extraordinary mechanism simultaneously fulfils the physical requirement for nucleating the growth of the helical particle and the biological requirement for specific recognition of the viral DNA.

Keywords: helix; protein disk; RNA; nucleation; X-ray analysis

1. EARLY HISTORY

Tobacco mosaic virus (TMV) is one of the simplest viruses known. Whether or not this simplicity of design is responsible for its success, its ready availability has made it convenient for biochemical and structural studies, and its simplicity has helped turn it, in one sense, into a paradigm for studies of biological structure and assembly. TMV has also provided an invaluable system for the development and application of new techniques in X-ray analysis and electron microscopy (EM), and also in biochemistry and genetics, although these are not our main concern here.

TMV has now been studied intensively for over 60 years (see table 1). Much of the early understanding of the properties of viruses came from the studies of TMV by W. M. Stanley and his colleagues, first at the Rockefeller Institute and then, in particular with H. C. Fraenkel-Conrat and R. C. Williams, at Berkeley, California, and by the group at Rothamsted and Cambridge, England (F. C. Bawden, N. W. Pirie and J. D. Bernal). The work at Berkeley and Rothamsted concentrated on the more biological and chemical properties of the virus. The complementary X-ray crystallographic study, started by Bernal at Cambridge before he moved to Birkbeck College, London, stopped during the the Second World War. It was taken up again by J. D. Watson, in a brief interlude from DNA, in Cambridge in 1952, by D. L. D. Caspar at Yale in 1954, and by Rosalind Franklin in 1953 at Birkbeck, where I joined her early in 1954. Investigations on the more chemical and physicochemical aspects of TMV were also started by G. Schramm in Tübingen during the War, and continued by himself, A. Gierer and H.-G. Wittmann. This early work (reviewed by Klug & Caspar 1960; Caspar 1963) led to an overall understanding of the helical subunit structure of TMV and the basis of its infectivity. Although subsequent investigations have led to a considerable refinement of this picture, at few points have its main features had to be changed. In this paper I will concentrate on the pathway of assembly of the virus from its constituent RNA and protein.

2. DESIGN AND ASSEMBLY

TMV is the classical example of a rod-shaped virus. Its rod shape results from its basic design, namely a regular helical array of identical protein subunits, in which framework is embedded a single molecule of RNA wound as a helix (figure 1*b*). It is, of course, the RNA that carries the genetic information, i.e. the capacity to instruct the host cell to make many copies of the virus. This general picture was already complete by 1958, when Rosalind Franklin died, and it seemed easy to comprehend how a structure of this type might be built out of identical subunits: the subunits might assemble themselves by repeated identical interactions, like adding steps in a spiral staircase, enclosing the RNA as a corkscrew-like thread as the rod extends. In other words, the assembly might be likened to growth at a screw dislocation in a crystal. We now know that this simple picture of assembly is wrong in all its essentials. The isolated protein can aggregate in different polymorphic forms, and Caspar (1963) suggested that these might play the role of intermediates in the assembly of the helical aggregate. The virus, in fact, assembles in a much more complex way, for what, with hindsight, we can see to be good physical and biological reasons. The story of how my colleagues and I came to suspect that the simple scheme was deficient and how the path of assembly was found has been told elsewhere (Klug 1979). I can only summarize the results here.

When the RNA and the coat protein of the virus are taken apart, the protein molecules alone under physiological conditions aggregate, not into a long helix, but into a 20S aggregate. We identified this as a 'disk' (figure 2)—a two-layer cylindrical structure, each layer consisting of a ring of 17 molecules, compared with the $16\frac{1}{3}$ molecules present in each turn of the assembled helix. The disk can be crystallized, but because of the large

 © 1999 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

Table 1. *Tobacco mosaic virus (TMV) structure and assembly*[a]*—a selected chronology 1936–1977*

1936–1939	Isolation, characterization, and first X-ray studies (Stanley, Bawden, Pirie, Bernal and Fankuchen)
1947–1955	Isolation of protein subunits and reaggregation into helical rods (Schramm)
1952–1955	Resumption of X-ray work—TMV shown to be a helix (Watson and Franklin)
1955	Self-assembly of infectious particles from separate components (Fraenkel-Conrat and Williams)
1956–1958	First one-dimensional Fourier maps, helical geometry and general description of structure (Franklin, Caspar, Klug and Holmes)

	virus		protein disk		assembly
	(Holmes, Klug; Cambridge)		(Leberman, Finch, Klug; Cambridge)	1958	physicochemical studies of the protein (Lauffer, Caspar)
1965	first 3D map; 12 Å resolution				
		1966	first X-ray studies, 17-fold symmetry	1963	Caspar: review on assembly
		1971	EM image reconstruction; 20 Å resolution	1970	(Klug, Butler; Cambridge) 'phase diagram' of protein aggregates (Durham)
1968	(Holmes, Stubbs; Heidelberg)	1972	first X-ray 3D map; 15 Å resolution (with Gilbert)	1971	20S disk shown to nucleate assembly (Butler)
1975	7 Å resolution	1975	5 Å resolution; chain traced (with Champness)	1976	nucleation region of RNA sequenced (Butler, Zimmern)
1977	4 Å resolution; RNA: protein contacts	1977	2.8 Å resolution; atomic model (with Bloomer)	1977	mechanisms of initiation and elongation shown
1978	(Stubbs, Namba, Caspar; Brandeis)				
1978→	higher resolution (Stubbs)				

[a]This deals mainly with the chronology of structural determinations and experiments on assembly. It omits the more biochemical and biological parts of the history, e.g. the discovery of the infectivity of the RNA by Gierer & Schramm and by Fraenkel-Conrat; the use of mutants in Melchers's laboratory to study the effects of changes in the protein subunit, and by Wittmann to test the genetic code; the sequencing of the protein in Tübingen and Berkeley.

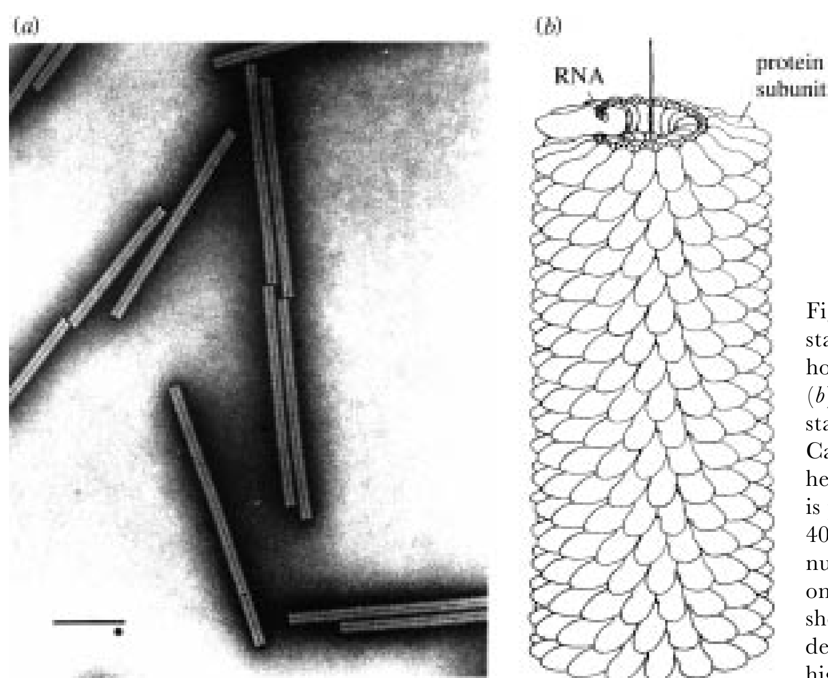

Figure 1. (*a*) Electron micrograph of negatively stained TMV particles (J. T. Finch). Note the hole down the middle. Scale bar, 100 nm. (*b*) Diagram summarizing the results of the first stage of structure analysis of TMV (Klug & Caspar 1960). The protein subunits form a tight helical array with $16\frac{1}{3}$ units per turn, and the RNA is packed between the turns at a radius of about 40 Å from the helix axis. There are three nucleotides per protein subunit. Only about one-sixth of the length of a complete particle is shown. The detailed X-ray structure was first determined by Stubbs *et al.* (1977), then carried to higher resolution by Namba & Stubbs (1986) and Namba *et al.* (1989).

molecular mass of the disk (600 000 Da) the determination of its exact structure by X-ray methods posed formidable technical and analytical problems. These were overcome and after a dozen years it was possible to construct an atomic model (figure 3) showing the detailed structure of the protein subunit and how it interacts with its neighbours (Bloomer *et al.* 1978).

This study was pursued to the end, not merely for the sake of knowing the structure of a protein subunit, but because we had shown earlier that the protein disk plays a crucial role in the assembly of the virus from its RNA and protein (Butler & Klug 1971). The disk combines with a specific initiation tract on the single-stranded viral RNA, which is located about

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

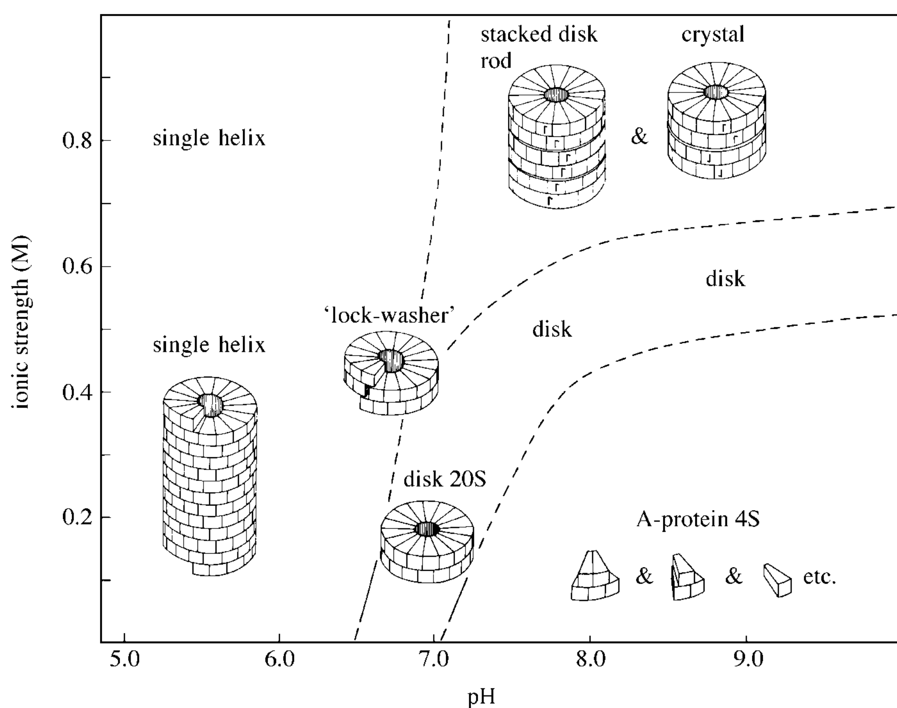

Figure 2. Diagram showing the ranges over which particular species of TMV protein participate significantly in the equilibrium (after Durham *et al*. 1971). This is not a conventional phase diagram: a boundary is drawn where a larger species becomes detectable and does not imply that the smaller species disappears sharply. The 'lock-washer' indicated on the boundary between the 20S disk and the helix is not well defined and represents a metastable transitory state observed when not enough time is allowed for the transition (fig. 1 in Butler, this issue). The boundaries are approximately correct for a protein concentration of 5 mg ml^{-1} at 20 °C, but not all species that have been observed are shown. The predominant species at pH 7.0 and $I=0.1$ is the '20S' aggregate (more accurately *ca*. 19.5S), which we identify as a two-layer polar disk. As the pH falls below pH 7, the sedimentation constant increases to values greater than 20S, characteristic of short helical aggregates (Butler *et al*. 1992). The stacked-disk structure is formed by prolonged incubation at alkaline pH, and represents an irreversible half-proteolysed aggregate (Durham 1972; Unwin & Klug 1974).

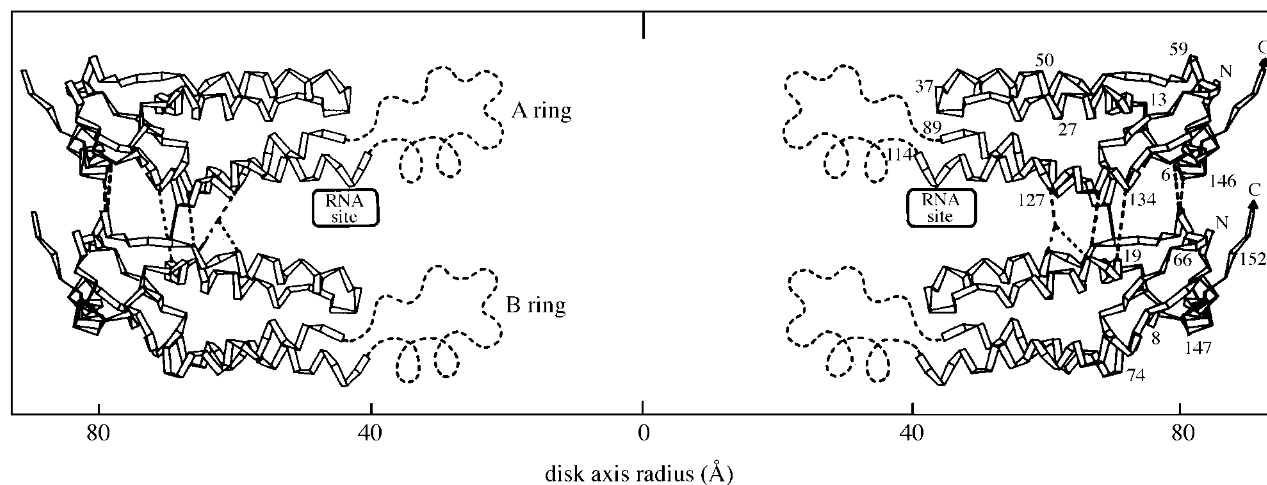

Figure 3. Section through a two-layer disk along its axis, reconstructed from the results of X-ray analysis to a resolution of 2.8 Å (Bloomer *et al*. 1978). The ribbons show the path of the polypeptide chain of the protein subunits. Subunits of the two rings can be seen touching over a small area toward the outside of the disk (where the side-chain contacts that occur between the rings have been drawn in as dashed and solid lines) but opening up into the 'jaws' toward the centre. The dashed lines at low radius indicate schematically the mobile portion of the subunits in the disk, extending in from near the RNA binding sites (between the rings *ca*. 40 Å radius as shown) to the edge of the central hole, which has a radius of 20 Å.

1000 bases from the 3′ end (Zimmern & Butler 1977; Zimmern & Wilson 1976), and then dislocates to begin helix growth. Foreign RNAs that do not have this tract are rejected.

When the sequence of bases of this initiation tract, or 'origin of assembly' as it is now called, was determined (Zimmern 1977), it was seen that it could be folded so that the initial binding site is exposed at the apex of a

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

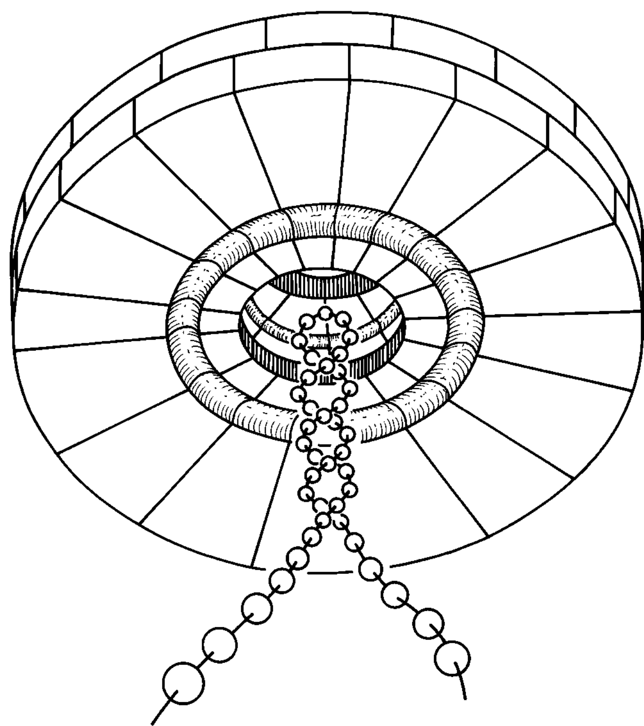

Figure 4. Nucleation of virus assembly is believed to occur by the insertion of a stem-loop hairpin structure of RNA into the central hole of the protein disk and between the two layers of subunits. This loop, formed by the nucleation region of the RNA origin of assembly, binds round the first turn, opening up the base-paired stem as it does so, and causes the disk to dislocate into a short helix. This 'closes the jaws', entrapping the RNA between the turns of protein subunits, and gives a start to the nucleoprotein helix (which can then elongate rapidly to some minimum stable size).

stem-loop structure. Now the X-ray studies had shown that the two layers of the protein disk are so arranged as to leave a gap between them, thus in effect extending the central hole of the disk, rather like a pair of 'jaws', as if waiting to 'bite' the RNA (figure 3). So we were able to develop the picture of initiation, or nucleation as a physicist would say, of assembly shown in figure 4 (Butler *et al.* 1976). The RNA hairpin loop inserts through the central hole of the disk, and the stem of the loop opens up and binds in the 'jaws' formed by the two layers of protein. The disk then dislocates into a helical structure, entrapping the RNA, after which elongation proceeds by the incorporation of further disks, pulling up more RNA through the central hole. There was some controversy surrounding this picture of growth after the initial stages, but there was no doubt about the role of the 20S aggregate (the disk) in initiating assembly, and also in further growth (see Butler, this issue). There has also been a controversy about the exact nature of the 20S aggregate, which we identified as a two-layer polar disk (see Appendix to this paper).

The disk is thus an obligatory intermediate in the assembly of the virus which, simultaneously, fulfils both the physical requirement for nucleating the growth of the helical particle and the biological requirement for specific recognition of the viral RNA. A most intricate structural mechanism has been evolved to give the process an efficiency and purposefulness, whose basis we now understand. TMV is self-assembling, self-nucleating and self-checking. The general conclusion derived from the story of TMV assembly is that one must distinguish between the design of a structure and the construction process used to achieve it. In the TMV structure all 2130 coat protein subunits (except the few at the ends of the particle) make the same non-covalent contacts with each other, and this specific bonding pattern repeated many times leads to a symmetrical final structure. There is nothing in the design of the completed structure which gives a hint that different bonding patterns, and non-equivalent ones at that, are required during the process of assembly.

The protein subunit under physiological conditions is designed not to form an endless helix, but a restricted two-layer variant of it—the disk—which is stable and which can be readily converted to the lock-washer, or helical, form. The disk therefore represents an intermediate subassembly by means of which the entropically difficult problem of nucleating helical growth is overcome. At the same time nucleation by the disk subassembly furnishes a mechanism for recognition of the homologous viral RNA (and rejection of foreign RNAs) by providing a long stretch of nucleotide binding sites for interaction with the special sequence of bases on the RNA. In this way specificity for the homologous viral RNA is achieved (see Butler, this issue).

APPENDIX A: THE IDENTITY OF THE 20S PROTEIN NUCLEATING AGGREGATE

The observations by Butler & Klug (1971) on the initiation of virus assembly *in vitro* at a pH of 7.0, or just above, established that the 20S protein aggregate nucleated the interaction of the virus coat protein specifically with the viral RNA to form the virus particle. The identification, by electron microscopy, of the 20S aggregate as the two-layer polar disk, found as the asymmetric unit of the protein crystal (figure 3), has been challenged by Caspar and his colleagues (Caspar & Namba 1990), who interpret it as a short protohelix. There is no direct way to visualize the packing arrangement in isolated 20S particles, but the original and later electron microscope observations (Durham *et al.* 1971; Sperling & Klug 1975, Butler *et al.* 1992) all point to a two-layer polar cylindrical structure as drawn in figure 2. Edge-on views are seen as two clear layers, and face-on views often show two overlapped circular structures, giving a 'figure of eight view', corresponding exactly to one of the interdisk packings in the protein crystal.

In further support of his argument, Caspar has recently studied the structure of some higher protein aggregates formed under different conditions having sedimentation constants of 38S, 45S and higher (Diaz-Avalos & Caspar 1998). They have termed these stacked disks, though the electron microscope image reconstruction shows them to be formed of dihedrally related stacked single layers. The biological relevance of the aggregates is doubtful. They are formed by dialysis for three days in 0.2 M ammonium sulphate at pH 8. Under such prolonged alkaline conditions we have found the coat protein to be prone to proteolysis (Durham 1972). In contrast, the standard 20S

Downloaded from rstb.royalsocietypublishing.org

disk preparations used by Butler for virus-particle reconstitution experiments are produced by dialysing the disaggregated protein (4S) in 0.1 M phosphate buffer at pH 7.0 for one day (cf. Durham & Klug 1971).

Finally, it should be said that the exact structure of the 20S aggregate, whether disk or protohelix, is not an issue fundamental to the picture of assembly described above. What cannot be doubted is that there exists a limited 20S aggregate of the coat protein that nucleates assembly, and that this same intermediate aggregate is also used in the elongation process towards the 5′ end of the RNA. What is at issue is the structural question of whether the switching from the nucleating aggregate to the helix is a quasi-equivalent or non-equivalent process (Caspar & Namba 1990).

REFERENCES

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. & 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., 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., 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.

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.

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

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

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

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

Klug, A. 1979 The assembly of tobacco mosaic virus: structure and specificity. *The Harvey Lecture* **74**, 141–162.

Klug, A. & Caspar, D. L. D. 1960 The structure of simple viruses. *Adv. Virus Res.* **13**, 1–63.

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 Visualisation 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.

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

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

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

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.

Downloaded from rstb.royalsocietypublishing.org