the last right column, without ever touching any nail in the diagonal of the lattice? Of course we can; and in a stupendous number of ways. Indeed if we knot our string around each column's nail selected at random, the odds are better than one-third that the resulting x will never equal the reported y [the exact probability being almost exactly $e^{-1} = (2.73\ldots)^{-1}$]. The 1928 Morgenstern point thus can still be a worry.

All this relates to rational expectationism, à la John Muth and others, as follows. A rational-expectation equilibrium time-profile of economic variables must be such that, if everyone were apprised of it, they would together all recreate exactly that profile. Hail to the Carnegie-Mellon workshops of the 1950's where Herbert Simon, John Nash, Abraham Charnes, William Cooper, John Muth, Charles Holt, Albert Ando, and Franco Modigliani made intellectual history with the perceptive support of Dean George Leland Bach.

Ad Hominem Matters

Hitler and Mussolini enriched American science. Along with Einstein, Weyl, Bethe, Ernst Mayr, von Neumann and so many others in the natural sciences, they presented us with such economists as Joseph Schumpeter, Wassily Leontief, Jacob Marschak, Gottfried Haberler, and Abraham Wald. Modigliani, by his youth, was at the end of this illustrious migration. By good luck, Jacob Marschak and Hans Neisser at the New School enabled him to land on his feet running. Great universities—Chicago, Illinois, Carnegie-Mellon, Northwestern, MIT—recognized his merits and he repaid their perspicuity. Every scholarly honor came his way, and fittingly early—presidencies of the American Economic Association, the Econometric Society, the American Finance Society, and so forth. Not only have governments benefited from his wisdom, but in addition he has helped universities and academies recognize undervaluations in Wall Street.

Still, there is one remarkable feature in Modigliani's scholarly profile. No lone scholar he; instead, dozens of his most famous contributions have been with joint authors, bearing such bylines as Modigliani-Ando, Modigliani-Brumberg, Modigliani-Grunberg, Modigliani-Miller, Modigliani-Samuelson, Modigliani-Drèze, and Modigliani-Papedemos. No one doubts Franco Modigliani's autonomous originality; all envy his ability to raise his own productivity and that of others by intense and joyful collaboration.

REFERENCES AND NOTES

1. F. Modigliani, *Econometrica* **12**, 45 (1944); reprinted in (*2*).
2. A. Abel, Ed., *The Collected Papers of Franco Modigliani* (MIT Press, Cambridge, MA, 1980).
3. F. Modigliani, *Rev. Econ. Stat.* **45**, 79 (1963); reprinted in (*2*).
4. S. Fischer, in *Macroeconomics and Finance: Essays in Honor of Franco Modigliani*, J. Bossons, R. Dornbusch, S. Fischer, Eds. (MIT Press, Cambridge, MA, in press).
5. F. Modigliani and R. Brumberg, in *Post Keynesian Economics*, K. K. Kurihara, Ed. (Rutgers Univ. Press, New Brunswick, NJ, 1954), pp. 388–436; reprinted in (*2*), along with other publications by Modigliani, and by Albert Ando and Franco Modigliani, on the life-cycle model.
6. F. Modigliani and M. H. Miller, *Am. Econ. Rev.* **48**, 261 (1958); M. H. Miller and F. Modigliani, *J. Business* **34**, 411 (1961); reprinted in (*2*).
7. For a sketch of competing paradigms in modern macroeconomics—post-Keynesian eclecticism, monetarism, and the new classical rational expectationism—see P. A. Samuelson and W. D. Nordhaus, *Economics* (McGraw-Hill, New York, 1985), part 3.
8. E. Grunberg and F. Modigliani, *J. Pol. Econ.* **62**, 465 (1954); reprinted in (*2*).
9. J. von Neumann and O. Morgenstern [*The Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, 1944), pp. 176–178] described the Holmes-Moriarty problem and referred to earlier work by Morgenstern.

Research Articles

Structure of Tobacco Mosaic Virus at 3.6 Å Resolution: Implications for Assembly

KEIICHI NAMBA AND GERALD STUBBS

X-ray fiber diffraction analysis of tobacco mosaic virus (TMV) has led to the building of a molecular model of the intact virus, based on a map at 3.6 Å resolution derived from five separated Bessel orders. This has been made possible by advances in the solution of the fiber diffraction phase problem. It is now possible to understand much of the chemical basis of TMV assembly, particularly in terms of intersubunit electrostatic interactions and RNA binding. Consideration of the molecular structure in conjunction with physical chemical studies by several groups of investigators suggests that the nucleating aggregate for initiation of TMV assembly is a short (about two turns) helix of protein subunits, probably inhibited from further polymerization in the absence of RNA by the disordering of a peptide loop near the inner surface of the virus.

TOBACCO MOSAIC VIRUS (TMV) HAS BEEN A MODEL SYSTEM for the study of protein–nucleic acid interactions and macromolecular assembly since Fraenkel-Conrat and Williams (*1*) showed that infective virus could be reconstituted from dissociated RNA and protein. Structural studies of the intact virus were begun by Bernal and Fankuchen (*2*), using x-ray fiber diffraction from oriented gels (*3*). These studies led to the calculation of a partially interpretable map at a nominal resolution of 4 Å (*4*). The virus is rod-shaped, 3000 Å long and 180 Å in diameter, with a central hole of diameter 40 Å. Approximately 2,130 identical protein subunits of molecular weight 17,500 form a helix of pitch 23 Å with 16⅓ subunits in every turn, protecting a single strand of RNA that follows the basic helix between the protein subunits at a radius of 40 Å. There are three nucleotides bound to each protein subunit.

Assembly of TMV is initiated by the binding of RNA to a

The authors are members of the Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235.

nucleating aggregate of protein having a sedimentation coefficient of about 20*S* (*5, 6*). Assembly then proceeds by the addition of 20*S* aggregates to the nucleoprotein complex (*5–7*), with smaller protein aggregates also being incorporated under appropriate conditions (*8*). The RNA is progressively inserted into its binding site from the inside of the growing virus particle, with the uncoated 5′ end of the RNA running down the central hole of the rod (*9, 10*).

TMV protein exists as a number of aggregates, depending on *p*H, ionic strength, and other factors (*7, 11*). One such aggregate, the 34-subunit disk (or 68-subunit four-layer disk), has been crystallized at *p*H 8 and high ionic strength (*12*), and its structure has been determined at 2.8 Å resolution (*13*). The disk has generally been assumed to be virtually identical to the nucleating aggregate, although spectroscopic and other studies indicate that there are definite differences between them (*14, 15*).

Structural studies of the intact virus are essential in order to study the interaction of protein and nucleic acid as well as the protein-protein interactions, which are very different in the disk and the virus (*16, 17*). Furthermore, a 25-residue loop of the protein is disordered in the disk and therefore not visible in the electron density map, but is ordered in the virus. Much of the RNA binding site and a major site responsible for control of assembly is in this loop (*4, 17*). The virus does not crystallize, and fiber diffraction methods have not until now been adequate to determine the structure in sufficient detail to understand the molecular mechanisms of assembly of TMV. However, recent developments in fiber diffraction analysis (*18–20*) have now enabled us to determine the structure of TMV at 3.6 Å resolution, to build a complete molecular model, and to consider the virus assembly in the light of this model.

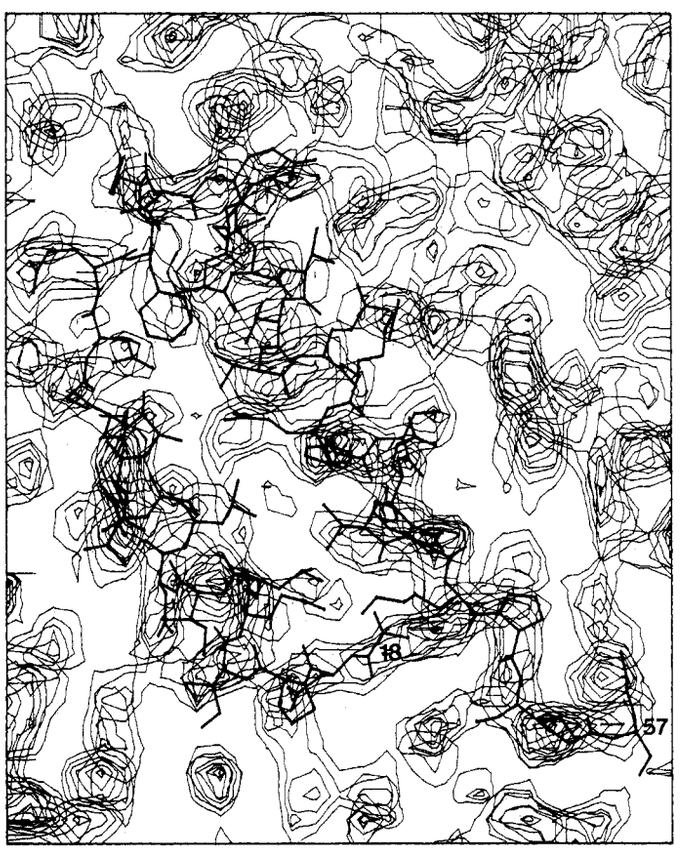


Fig. 1. The electron density map after isomorphous replacement with layer line splitting and solvent-flattening refinement. Five sections at an angle of 75° to the particle axis are shown, together with the refined molecular model between residues 18 and 57 (the left and right slewed helices). Diffraction patterns were recorded on film, digitized, converted from film space to reciprocal space (*21*), and corrected for nonlinear response and geometric factors. The angular deconvolution method (*18*) was used to obtain reliable estimates of intensities and layer line positions. Up to five G terms were separated and phased for each data point, using intensities from six derivatives and layer line splitting from four. Phases were refined against intensities by alternating a solvent-flattening procedure with two-dimensional (that is, conventional crystallographic) isomorphous replacement, in which the G term separation from the previous step was retained, and new phases were determined (*20*). A map of 3.6 Å resolution calculated in this way was used with an Evans and Sutherland computer graphics system and the program FRODO (*26*) to build a model. The model was refined iteratively with the phases, using the model to calculate the separation of the G terms, and calculating new phases by two-dimensional isomorphous replacement. This procedure converged after five cycles.

Structure analysis. Fiber diffraction data were collected for TMV and the six heavy atom derivatives used by Stubbs, Warren, and Holmes (*4*). The data analysis has been described (*20*). Because fiber diffraction data are cylindrically averaged, the diffracted intensity (*22, 23*) is

$$I = \sum \mathbf{G}\mathbf{G}^*$$

where **G** is a complex Fourier-Bessel structure factor (*24*), analogous to the crystallographic **F**. Both phases and magnitudes of **G** terms must be determined in order to calculate an electron density map; thus the phase problem in fiber diffraction is multidimensional. The number of significant terms in this equation depends on the radius and symmetry of the diffracting particle and on the resolution; for TMV at 3.6 Å resolution this number can be up to five. A multidimensional analog of crystallographic isomorphous replacement (*25*) can be used to separate and phase these terms, but it requires a large number of heavy-atom derivatives: two for each term. Measurement of the fine splitting of the layer lines (stemming from the fact that the TMV helix does not repeat perfectly, but has 49.02 subunits in three turns) provides further phase information (*19*) and can reduce the number of heavy-atom derivatives required by up to a factor of 2. Initial phases were calculated in this way, and refined by solvent flattening or other methods (*20*). Part of the final map is shown in Fig. 1, together with the model, further refined by the least-squares procedure with stereochemical restraints (*27*). During this refinement the *R* factor (*28*) of the model was reduced from 0.31 to 0.14. The adaptation of restrained least-squares refinement to fiber diffraction has been described (*29*); but some indication of the quality of the refinement is given in Fig. 2, where refined temperature factors of the main chain atoms are compared with those of the protein in the disk crystals (*30*). The complete set of atomic coordinates will be deposited in the Protein Data Bank at Brookhaven National Laboratory when further refinement and extension of resolution has been completed.

The protein subunit. The backbone structure of the protein subunit in the virus (Fig. 3A) is quite similar to that of the protein disk aggregate (*13*) (Fig. 3B). The central part of the molecule consists of four closely packed α-helices which form a hydrophobic core extending radially from 45 Å to 65 Å. These are referred to as the left and right slewed helices (the top layer of each subunit in Fig. 3A, toward and away from the reader, respectively), and the left and right radial helices; LS, RS, LR, and RR (*16*). They include residues 19 to 32, 38 to 48, 111 to 135, and 73 to 87. There are three other short helical segments: one near the NH_2-terminus, one near the COOH-terminus, and a very short segment (V) running approximately vertically for about 1½ turns from Pro^{102}, at the inner surface of the virus.

At the outer end of the four core helices, at a radius of about 70 Å, is a small region of β structure, connecting the helices. Outside this is a cluster of six aromatic residues from the NH_2- and COOH-terminal sections of the protein and from a loop connecting RS and RR. This loop, in the top layer of the subunit, interacts with the

COOH-terminal helix in the bottom layer through two salt bridges on the outer surface of the virus. The COOH-terminal residues 155 to 158, which are presumed to be somewhat disordered, have not yet been located in the map.

The innermost loop of the peptide chain, residues 90 to 113, is not visible in the electron density map from the disk crystal because of disorder (*13*), but is clearly recognizable in the virus. The chain extends inward from the end of RR at residue 87, forming two phosphate binding sites for the RNA strand on the top surface of the subunit and part of one of the base binding sites for the RNA strand below. The chain continues to the top of an approximately vertical segment which includes V, turning downward to form a β bend at residues 97 to 100. At Thr^{107} the chain turns out again, connecting to LR at Thr^{111}. LR forms most of the RNA base binding sites and also interacts with the RNA backbone.

This chain folding, in particular the high proportion of secondary structure, suggests that the molecule is quite rigid, and indeed, a rigid body transformation between the alpha carbon coordinates of the virus and unrefined coordinates for the disk protein (omitting the loop disordered in the disk) leads to a root-mean-square (rms) difference of only 0.7 Å, although there are significant side-chain rearrangements.

Subunit packing. The protein-protein interactions in the virus may be considered in three classes: a major class of top-to-bottom contacts between a subunit and the subunit 16 units further along the basic helix (that is, the next subunit along the 16-start left-handed helix), a minor class of top-to-bottom contacts along the 17-start right-handed helix, and the class of side-to-side contacts. The top-to-bottom contacts are quite different from those found in the disk: whereas in the virus, each subunit is displaced about a third of a subunit to the left (viewed from outside) of its lower neighbor, the displacement in the disk is about a fifth of a subunit to the right (*16*). Furthermore, the axial tilt of the subunits is very different (Fig. 3). As a result of this difference (12° for the top layer of the disk, and 22.5° for the bottom layer), the high-radius top-to-bottom contacts in the disk are completely absent in the virus, leaving the large, flat (1-start) helical groove that has long been known to be characteristic of TMV (*31*). At radii less than about 70 Å, the top-to-bottom contacts are very close in the virus, consisting of close packing between the α-helices and interactions between the inner loops. There is a continuous cavity along the 1-start helix, which accommodates the RNA. Most of the top-to-bottom contacts are between the subunits related by the 16-start helices, but the minor 17-start contact is of some interest, in that it includes an ion pair (Glu^{50} and Arg^{134}) which forms part of a major intersubunit charge interaction (see below).

The side-to-side packing is closely related to that found in the disk; but, even here, there are significant differences, as is shown by a rigid-body transformation superimposing 1 subunit of the virus helix on a subunit of the top layer of the disk. Examining the effect of this transformation on the azimuthal neighbors of the superimposed subunits (Fig. 4), we find a radial displacement of about 1 Å, a twist of about 2.5° about a horizontal axis at right angles to and near the outer end of LR, and a hinge movement of about 4.5° about a radial line roughly in the LS helix. This difference in the hinge angle between adjacent subunits is reflected in a change of tilt of the subunits relative to the viral axis, and thus gives the virus a convex end compared with the relatively flat top of the disk. The 2.5° twist is primarily responsible for the helical quaternary structure of the virus. The overall effect includes an even closer side-to-side packing between subunits in the virus than is seen in the disk. The interacting residues are generally the same in the virus and the disk, with the different packing accommodated by large side-chain conformational changes.

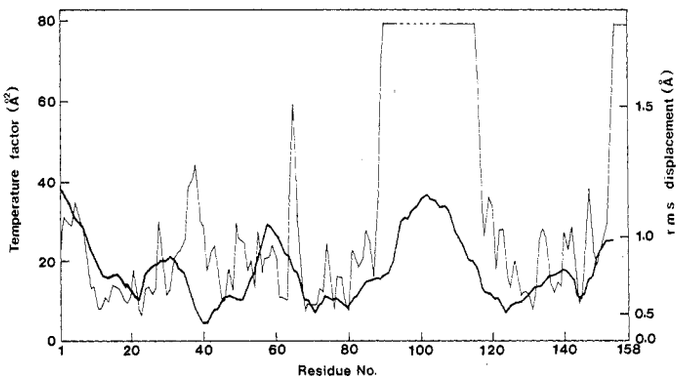


Fig. 2. The mean temperature factor of the main-chain atoms in each residue plotted against residue number for the coat protein of TMV. The heavy line is for the protein in the virus, as refined in this work, while the light line is for the protein in the crystalline disk (*30*). The virus curve is smoother than that of the protein disk, but this merely reflects different types of restraint used for the temperature factor refinements. Apart from this feature, the curves show considerable similarity. The rms displacement is calculated from $U = 3B/8\pi^2$, where U is the displacement and B is the temperature factor. The peptide chain around residues 95 to 110 in the inner loop, which is disordered in the protein disk, also has relatively high temperature factors in the virus.

The highest peak in the radial density distribution (*32, 33*) is at a radius of about 23 Å. This region is occupied by the vertical chains containing the V helices. These chains pack closely together, each filling a space 9 Å wide and 23 Å high (*4*) and forming a dense wall around the central hole of the virus particle.

The subunit packing has been displayed with the use of computer graphics (see figure 2 in *34*).

Intersubunit charge interactions. There are two major networks of intersubunit charge interactions in the virus particle. The first is at low radius, apparently including all the charged residues near or inside the RNA radius except Glu^{97}, and is in fact continuous with the charge interactions that form the phosphate binding sites for the RNA. This has been referred to as the "carboxyl cage" (*4, 17*). Both lead and uranyl ions bind to a site in this region near the inner surface of the virus, probably coordinated with Glu^{106} from one subunit and Glu^{95}, Asp^{109}, and Asn^{101} from its side-to-side neighbor. The site may also include main-chain carbonyl groups. The metal binding and the appearance of this site suggest that it is one of the two calcium binding sites identified in titration experiments (*35, 36*), and the site of one of the anomalously titrating protons found in TMV (*37–39*). Caspar (*37*) predicted that these protons would be bound to pairs of carboxyl groups, and the pair Glu^{95}–Glu^{106} is the most likely candidate, since Asp^{109} is not strongly conserved among TMV strains. It is not possible to make this assignment unambiguously, since the diffracting samples were at *p*H 8 or higher, with all the carboxyl groups ionized, and the chain folding stabilized by the RNA binding alone. One might possibly speculate that the relatively high temperature factor in this area (Fig. 2) is due to this ionization.

Still in this charge network, but at higher radius, there are two side-to-side intersubunit salt bridges: Arg^{113}–Asp^{115}, and Asp^{88}–Arg^{122}, with a base binding site between them. The first of these is not seen in the protein disk because Arg^{113} is in the disordered inner loop in that structure. The second is seen in the disk (*13*), although the difference in subunit packing necessitates a considerable change in the conformation of the side chains. Asp^{115}, Asp^{116}, and the phosphate groups form a concentration of negative charge. There does not appear to be any direct interaction between Asp^{115} and Asp^{116}, but Asp^{116} is extremely close to one of the phosphate groups.

The second charge interaction is located between 55 Å and 60 Å

in radius, around the outer end of the α-helical core, at the interface of four subunits: two each from two consecutive turns of the basic (1-start) helix. A cylindrical section of the four subunits, seen from the virus axis, is shown in Fig. 5A. Four α-helices, one from each subunit, contribute side chains to the interaction. Asp^{77} from the top right subunit forms a salt bridge with Arg^{71} of the same subunit, and is also very close (less than 4 Å in this model) to Glu^{50} from the bottom right subunit. This Glu^{50} forms an ion pair with Arg^{134} from the top left subunit, as does Arg^{134} with Glu^{131} from the same subunit. It appears that Asn^{29} from the bottom left subunit may form hydrogen bonds with both Arg^{134} and Glu^{131}, and that hydrogen bonds may also be formed between Arg^{71} and Thr^{81} within a subunit and between Tyr^{72} and Thr^{28} across the 1-start helix intersubunit boundary. These intricate and specific interactions contribute to the helical packing of the subunits in the virus.

The carboxyl pair Asp^{77}–Glu^{50} forms the second metal-binding site in the uranyl fluoride derivative (perhaps with Asn^{126}), suggesting that this is the second site of calcium binding (*35, 36*) and anomalously titrating protons (*37–39*). Because of the very different packing arrangement of subunits, none of the intersubunit charge

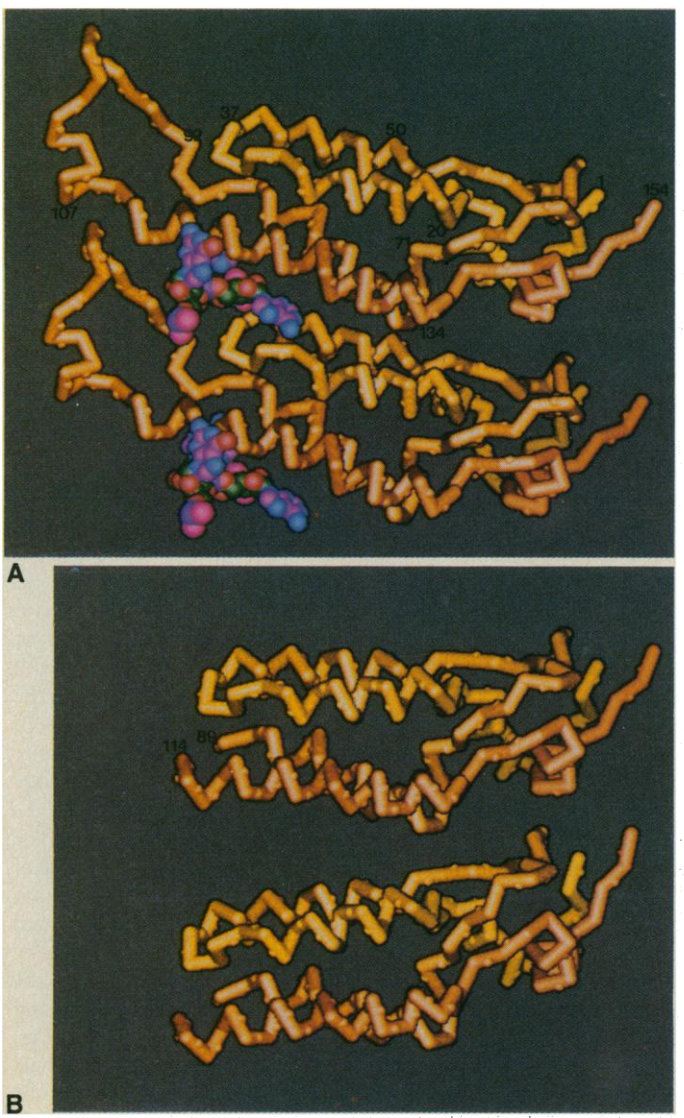


Fig. 3. The folding of the polypeptide chain and the axial packing of subunits in the virus and the protein disk. (A) Two subunits of the virus related by the 16-start helix, viewed perpendicular to the axis. The virus axis runs vertically, to the left of the figure, and the outer surface is on the right. (B) The same view of two subunits from the top and bottom layers of the protein disk. The disordered residues 90 to 113 are not included. The alpha-carbon atoms are connected by cylinders 2 Å thick and color-coded for sequence, from the NH_2-terminus (yellow) to the COOH-terminus (brown). The direction of each main chain carbonyl group is indicated by a small bump on the cylinder. The atoms of the RNA are color-coded as follows: uncharged oxygen, red; charged oxygen, crimson; uncharged nitrogen, blue; sugar carbon, green; base carbon, purple; phosphorus, blue-purple. The diameters of the atoms are 3.5 Å for phosphorus, 3 Å for charged oxygen, and 2.5 Å for all others. The graphics scheme used in (A) and (B) and in Fig. 5, A and B was developed in collaboration with D. L. D. Caspar from the system described in (*34*).

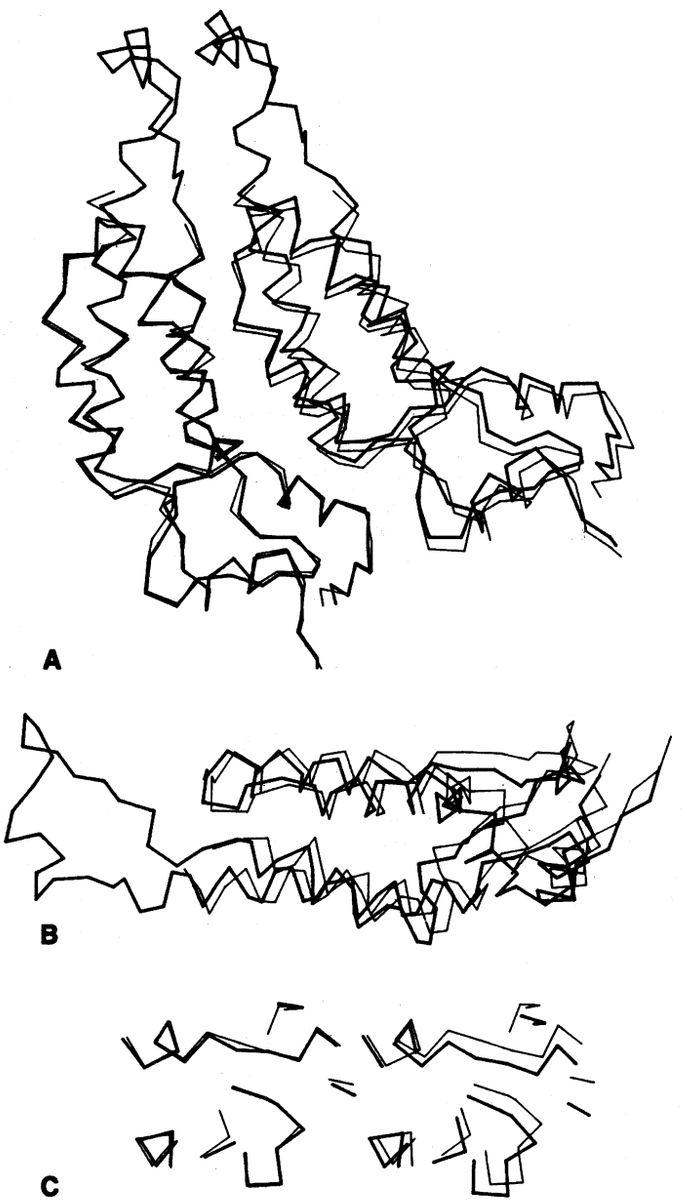
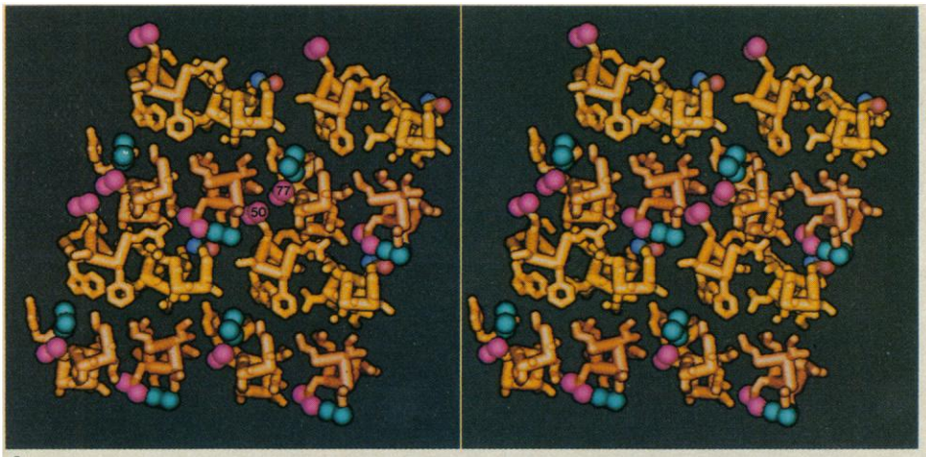
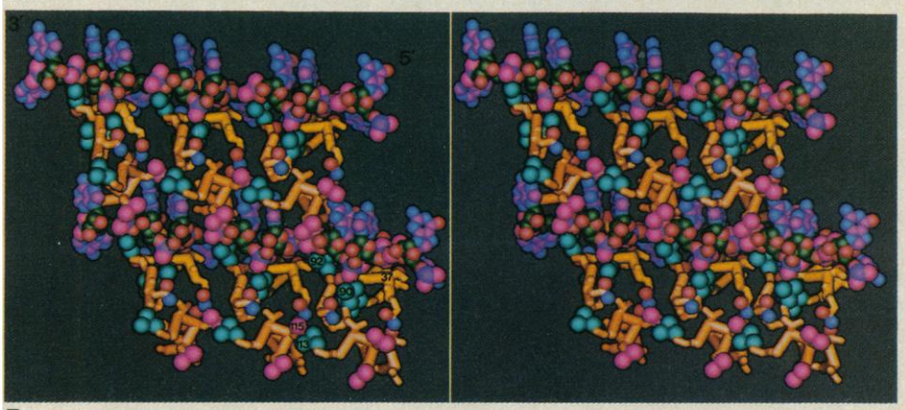


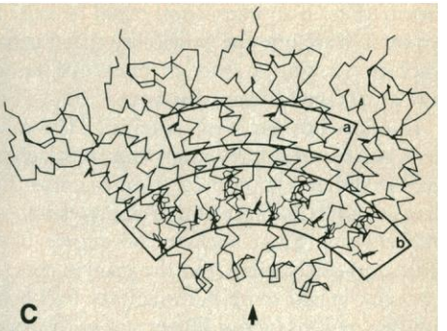
Fig. 4. Comparison of the subunit packing in the virus and the protein disk. Heavy line: virus; light line: disk. (A) Top view of two subunits of the disk (top layer) with two subunits of the virus. The virus subunits have been subjected to a rigid body transformation so that the virus subunit on the left is superimposed upon the corresponding disk subunit. There is consequently a radial displacement of about 1 Å between the subunits on the right. Because of the transformation, the virus axis is not perpendicular to the plane. (B) Azimuthal projection of the subunits on the right in (A). There is a 2.5° twist about a point near the outer end of the left radial helix. (C) Cylindrical sections of the subunits in (A) at radii between 65 Å and 72 Å. There is a difference of about 4.5° in the hinge angle about the left slewed helix.



A



B



C

Fig. 5. (A) Stereo pair of a cylindrical section through four subunits in the virus, viewed from the axis, showing one of the intersubunit electrostatic interactions. Residues 20 to 29, 47 to 53, 71 to 80, and 128 to 135 are included. The alpha-carbon backbone is represented as in Fig. 3, with side-chain atoms connected by thin rods. The groups involved in the electrostatic interaction (from top right to bottom left in the center of the figure) are Arg^{71}, Asp^{77}, Glu^{50}, Arg^{134}, Glu^{131}, and Asn^{29}. These are shown as spheres, color coded as follows: chargeable oxygen, crimson; unchargeable oxygen, red; chargeable nitrogen, cyan; unchargeable nitrogen, blue. The diameters of these atoms are 3 Å for chargeable oxygen and nitrogen, and 2.5 Å for unchargeable atoms. (B) Stereo pair of a cylindrical section through the RNA binding site, viewed from the axis. Three protein subunits and twelve nucleotides from each of two turns of the virus helix are shown. The protein includes residues 36 to 41 at the hairpin loop connecting the left and right slewed helices, 90 to 92 near the end of the right radial helix and 113 to 120 in the left radial helix. Two bases extend vertically into the intersubunit space between two left radial helices and one base extends along this helix (see Fig. 3A). Asp^{115} from the left and Arg^{113} from the right form an intersubunit bridge just in front of the two vertical bases. (C) Line drawing of four subunits of the virus, viewed from the top, showing the sections taken to produce (A) and (B).

interactions in this system can occur in the protein disk, and these two carboxyl groups are in fact 11 Å apart there. The significance of this in explaining the characteristic titration behavior of the protein is discussed below.

Protein–RNA interactions. The RNA structure has been described by Stubbs and Stauffacher (*40*), whose model was based on an electron density map at 4 Å resolution (*4*). The structure found here is very similar to that model, the root-mean-square difference between the two being only 0.7 Å. Figure 5B is a cylindrical section of the virus viewed from the axis, showing the interactions of the RNA with the protein. The nucleotides are numbered from the 5′ end to the 3′ end of the RNA. The phosphate groups are neutralized by Arg^{41}, Arg^{90}, and Arg^{92}, but they do not all form simple ion pairs. Phosphate group 1, at the lowest z coordinate, is neutralized by Arg^{90}, and phosphate group 2, about 4 Å higher in z, is neutralized by Arg^{92}. The main protein chain rises steeply between Arg^{90} and Arg^{92} (see Fig. 3A). Phosphate group 3 does not appear to be directly neutralized by a positive charge, but forms a hydrogen bond with Thr^{37} in the hairpin loop connecting LS and RS. Arg^{41} extends beneath this loop toward phosphate group 1, but does not approach it as closely as Arg^{90} does. Asp^{116} from the protein subunit above the RNA is close to phosphate group 2. This concentration of negative charge may be partially neutralized by Arg^{112}, also from the subunit above, but the electron density for this side chain is not clear.

All three bases lie flat against the LR helix. Base 1 presents its hydrophobic surface to a methyl group from Val^{119}, while base 3 is close to the α-helical main chain between Asp^{116} and Ala^{117}. These two bases stack together and point up into a cavity formed by the LR helix, the extended chain following the RR helix, the LR helix of the 3′ neighboring subunit, and the intersubunit salt bridges Arg^{113}–Asp^{115} and Arg^{122}–Asp^{88}. Base 2 lies along the LR helix, between this helix and the connecting LS-RS loop from the subunit below (Fig. 3A). Ser^{123} and Asn^{127} from the top subunit and Asn^{33} and Glu^{34} from the bottom subunit provide a hydrophilic environment for the polar parts of the base. Built as purine bases, base 1 and base 2 are in the *anti* conformation and base 3 is in the *syn* conformation. The electron density for base 2 indicates another possible conformation, still *anti*, lying flat against the LR helix, but pointing toward the 5′ end of the RNA instead of the 3′.

The interactions discussed above are not specific to the RNA sequence. It is possible to postulate the existence of hydrogen bonds between the RNA bases and the protein, and some of these bonds could enable the different base-binding sites to bind certain bases preferentially, although such specificity must remain tentative at present. If base 1 is guanine, it could interact with the Arg^{122}–Asp^{88} and Asp^{115}–Arg^{113} intersubunit salt bridges, atom O6 forming a hydrogen bond with Arg^{122}, and atom N2 with Asp^{115}. If base 3 is adenine, the main chain carbonyl groups of Ala^{86} and Thr^{89} point toward N6, suggesting hydrogen bonds. The amide group of Asn^{127} is placed so that it could form hydrogen bonds with base 2; if the base is adenine, it would form bonds with atoms N1 and N6. Because of the capacity of the amide group to rotate about the β-γ carbon (C_β-C_γ) bond, this need not be a base-specific interaction. However, if the idealized planar base-amide complex were to be

distorted, such an interaction could become base specific. Although we cannot exclude the possibility that further refinement at higher resolution might produce changes, our model is distorted in such a way as to favor adenine binding.

Implications for viral assembly. The model of TMV presented here, made possible by improved approaches to the phase problem in fiber diffraction, permits us to examine the structure in sufficient detail to address several important questions relating to the assembly of the virus. These include the degree of similarity between the structures of the virus and the protein disk. The two structures are by no means the same. Although the backbone structures are similar in those regions (about 70 percent of the total structure) where they show a similar degree of order, there are major side-chain differences; not only in the top-to-bottom protein interfaces (where the grossly different subunit packing requires such differences), but also in the side-to-side contacts. This is because the relation between the two structures is more complicated than the simplest possible "opening up" of a layer of the closed disk to form the gently sloping 1-start helix of the virus.

The binding sites in the uranyl fluoride heavy-atom derivative appear to be the most likely candidates for the sites of the anomalously titrating carboxyl groups postulated by Caspar (*37*). These groups play an essential part in the assembly and disassembly of the virus particle: the electrostatic repulsion of the carboxylate groups, counteracted by the binding of the RNA, as well as cation binding under appropriate conditions, provides a sensitive switch, active under physiological conditions, to control the state of aggregation of the protein (*4, 37, 41*). The two sites that we found are both intersubunit pairs of carboxyl groups; Glu^{95}–Glu^{106}, at 25 Å radius, is in the side-to-side protein interface, while Glu^{50}–Asp^{77}, at 58 Å radius, is in the top-to-bottom interface. The positions of these sites correlate well with the titration results of Shalaby and Lauffer (*39*), who found that at *p*H 7, 20°C, and low ionic strength, where the protein exists largely as the 20*S* aggregate, half a proton is bound to each protein monomer. This suggested a binding site between the layers of the aggregate, such as our 58 Å radius site. As noted above, however, no such site exists in the protein disk (*13*). This suggests that at neutral *p*H and low ionic strength, the normal conditions for viral assembly, the 20*S* nucleating aggregate is not the disk, but a helical aggregate of about the same size, that is, about two turns of helix. Nonetheless, it shares many properties with the disk, in particular, as shown by nuclear magnetic resonance (NMR) studies (*42*), the disordered inner loop. This identification of the 20*S* aggregate is in agreement with the conclusions of Correia *et al.* (*43*), who have shown by sedimentation equilibrium experiments that the number of subunits in the aggregate is 39 ± 2, and with circular dichroism experiments (*14, 15*), which show that the protein conformation under nucleating conditions is different from that at higher *p*H and ionic strength.

It appears that the two-turn helix may be sterically prevented from growing longer by the disordering of the inner loop. Lowering the *p*H protonates the carboxyl groups, removing the electrostatic repulsion that causes the disorder, and permitting the inner loop to fold and a long helical aggregate to form. Under physiological conditions, however, the specific binding of RNA is needed to induce the disordered loop to fold. This in turn allows elongation to take place and assembly to proceed.

If the short helix of two turns or a little more is considered to be the nucleating aggregate, initial binding of the RNA is unlikely to be between the turns. One might expect the array of arginine groups on the top surface of the aggregate to bind RNA phosphate groups nonspecifically. When the specific AAG-rich binding sequence at which assembly is initiated (*44*) is encountered, the high affinity of the protein for this sequence—found by Steckert and Schuster (*45*) and supported by our evidence for specific binding sites—could lead to further binding of the RNA, either by intercalation or by bases binding to the bottom surface of the aggregate. Although this description of the molecular basis for initiation of viral assembly differs in some details from those of Butler *et al.* (*9*) and Lebeurier *et al.* (*10*), it is essentially consistent with their observations and the proposed doubling back of the RNA in their mechanism of assembly. It is particularly easy to visualize elongation of the growing viral rod by addition of short helices, since the top surface of the short helix is complementary to the bottom surface of the rod.

REFERENCES AND NOTES

1. H. Fraenkel-Conrat and R. C. Williams, *Proc. Natl. Acad. Sci. U.S.A.* **41**, 690 (1955).
2. J. D. Bernal and I. Fankuchen, *J. Gen. Physiol.* **28**, 111 (1941).
3. J. Gregory and K. C. Holmes, *J. Mol. Biol.* **13**, 796 (1965).
4. G. Stubbs, S. Warren, K. Holmes, *Nature (London)* **267**, 216 (1977).
5. P. J. G. Butler and A. Klug, *Nature (London) New Biol.* **229**, 47 (1971).
6. T. Ohno, R. Yamaura, K. Kuriyama, H. Inoue, Y. Okada, *Virology* **50**, 76 (1972).
7. P. J. G. Butler and A. C. H. Durham, *Adv. Protein Chem.* **31**, 187 (1972).
8. S. J. Shire, J. J. Steckert, T. M. Schuster, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 256 (1981).
9. P. J. G. Butler, J. T. Finch, D. Zimmern, *Nature (London)* **265**, 217 (1977).
10. G. Lebeurier, A. Nicolaieff, K. E. Richards, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 149 (1977).
11. A. C. H. Durham, J. T. Finch, A. Klug, *Nature (London) New Biol.* **229**, 37 (1971).
12. R. Leberman, J. T. Finch, P. F. C. Gilbert, J. Witz, A. Klug, *J. Mol. Biol.* **86**, 179 (1974).
13. A. C. Bloomer, J. N. Champness, G. Bricogne, R. Staden, A. Klug, *Nature (London)* **276**, 362 (1978).
14. D. Vogel and R. Jaenicke, *Eur. J. Biochem.* **61**, 423 (1976).
15. K. Raghavendra, M. L. Adams, T. M. Schuster, *Biochemistry* **24**, 3298 (1985).
16. J. N. Champness, A. C. Bloomer, G. Bricogne, P. J. G. Butler, A. Klug, *Nature (London)* **259**, 20 (1976).
17. G. Stubbs, in *Biological Macromolecules and Assemblies*, volume 1, *Virus Structures*, F. A. Jurnak and A. McPherson, Eds. (Wiley, New York, 1984), p. 149.
18. L. Makowski, *J. Appl. Crystallogr.* **11**, 273 (1978).
19. G. Stubbs and L. Makowski, *Acta Crystallogr. Sect. A* **38**, 417 (1982).
20. K. Namba and G. Stubbs, *ibid. A* **41**, 252 (1985).
21. R. D. B. Fraser, T. P. Macrae, A. Miller, R. J. Rowlands, *J. Appl. Crystallogr.* **9**, 81 (1976).
22. J. Waser, *Acta Crystallogr.* **8**, 142 (1955).
23. R. E. Franklin and A. Klug, *ibid.*, p. 777.
24. A. Klug, F. H. C. Crick, H. W. Wyckoff, *ibid.* **11**, 199 (1958).
25. G. J. Stubbs and R. Diamond, *Acta Crystallogr. Sect. A* **31**, 709 (1975).
26. T. A. Jones, in *Computational Crystallography*, D. Sayre, Ed. (Oxford Univ. Press, Oxford, 1982), p. 300.
27. W. A. Hendrickson and J. H. Konnert, in *Computing in Crystallography*, R. Diamond, S. Ramaseshan, K. Venkatesan, Eds. (Indian Academy of Science, Bangalore, 1980), p. 13.01.
28. The *R* factor is analogous to the usual crystallographic *R* factor, being the sum over all observations of the differences between observed and calculated structure factors (VI), divided by the sum of the observed structure factors.
29. G. Stubbs, K. Namba, L. Makowski, *Bioplys. J.*, **49**, 58 (1986).
30. E. Westhof, D. Altschuh, D. Moras, A. C. Bloomer, A. Mondragon, A. Klug, M. H. V. Van Regenmortel, *Nature (London)* **311**, 123 (1984).
31. R. E. Franklin and A. Klug, *Biochim. Bioplys. Acta* **19**, 403 (1956).
32. D. L. D. Caspar, *Nature (London)* **177**, 928 (1956).
33. R. E. Franklin, *ibid.*, p. 928.
34. K. Namba, D. L. D. Caspar, G. J. Stubbs, *Science* **227**, 773 (1985).
35. W. H. Gallagher and M. A. Lauffer, *J. Mol. Biol.* **170**, 905 (1983).
36. ———, *ibid.*, p. 921.
37. D. L. D. Caspar, *Adv. Protein Chem.* **18**, 37 (1963).
38. H. Fraenkel-Conrat and K. Narita, in *Symposium on Protein Structure*, A. Neuberger, Ed. (Wiley, New York, 1958), p. 249.
39. R. A. F. Shalaby and M. A. Lauffer, *J. Mol. Biol.* **116**, 709 (1977).
40. G. Stubbs and C. Stauffacher, *ibid.* **152**, 387 (1981).
41. D. L. D. Caspar, in *Structure-Function Relationships of Proteins*, R. Markham and R. W. Horne, Eds. (North Holland, Amsterdam, 1976).
42. O. Jardetsky, K. Akasaka, D. Vogel, S. Morris, K. C. Holmes, *Nature (London)* **273**, 564 (1978).
43. J. J. Correia, S. Shire, D. A. Yphantis, T. M. Schuster, *Biochemistry* **24**, 3292 (1985).
44. D. Zimmern, *Cell* **11**, 463 (1977).
45. J. J. Steckert and T. M. Schuster, *Nature (London)* **299**, 32 (1982).

46. We thank D. L. D. Caspar for advice and encouragement, and for discussions concerning TMV assembly and the nature of the nucleating aggregate; K. C. Holmes for guidance during the early stages of this work; and A. C. Bloomer for providing a preliminary set of atomic coordinates for the protein disk. Computer model building facilities were provided by Columbia University and EMBL, Heidelberg, and some of the computations were carried out with the use of a VAX 11/780 computer at Brandeis University provided by NIH grant 3-R01-GM21189-09S1 (D. J. DeRosier). Supported by NIH grants GM25236 and GM33265.

19 September 1985; accepted 28 January 1986