Journal of Supramolecular Structure 12:305–320 (1979)
Biological Recognition and Assembly 155–170

Protein-RNA Interactions During TMV Assembly

K. C. Holmes

Max Planck Institut für medizinische Forschung, Heidelberg, Germany

A review of the structural studies of tobacco mosaic virus (TMV) is given. TMV is essentially a flat helical microcrystal with 16 1/3 subunits per turn. A single strand of RNA runs along the helix and is deeply embedded in the protein. The virus particles form oriented gels from which high-resolution X-ray fiber diffraction data can be obtained. This may be interpreted by the use of six heavy-atom derivatives to give an electron density map at 0.4 nm resolution from which the RNA configuration and the form of the inner part of the protein subunit may be determined. In addition, the protein subunits form a stable 17-fold two-layered disk which is involved in virus assembly and which crystallizes. By the use of noncrystallographic symmetry and a single heavy-atom derivative, it has been possible to solve the structure of the double disk to 0.28 nm resolution. In this structure one sees that an important structural role is played by four alpha-helices, one of which (the LR helix) appears to form the main binding site for the RNA. The main components of the binding site appear to be hydrophobic interactions with the bases, hydrogen bonds between aspartate groups and the sugars, and arginine salt bridges to the phosphate groups. The binding site is between two turns of the virus helix or between the turns of the double disk. In the disk, the region proximal to the RNA binding site is in a random coil until the RNA binds, whereupon the 24 residues involved build a well-defined structure, thereby encapsulating the RNA.

Key words: tobacco mosaic virus, structure, RNA-binding site, assembly, protein-nucleic acid interactions

There are still very few systems from which we can obtain detailed information about the geometry of the interactions between proteins and nucleic acids. One such is tobacco mosaic virus (TMV). TMV particles are rod-like, 300 nm long, and 18 nm in diameter. TMV consists of 2,140 protein subunits, each of molecular weight 17,420 daltons (158 residues), arranged on a helix of pitch 2.3 nm with 16 1/3 subunits per turn. Winding through this helix is a single strand of RNA 6,400 nucleotides long (Fig. 1) [1, 2].

Structural studies on TMV have a long history. In 1936 Bernal obtained the first X-ray diffraction patterns from orientated gels of TMV and realized the implication of the existence of layer-lines in this pattern, namely, that the particle is built up from a regular arrangement of repeating subunits [3]. At the time, this conclusion did not find immediate

Received May 4, 1979; accepted August 10, 1979.

0091-7419/79/1203-0305$02.90

acceptance. Furthermore, the detailed interpretation of the diffraction pattern eluded Bernal and Fankuchen. Only when the necessary theoretical development, the elucidation of the structure factor of a helix, had been carried out in 1952 [4] was it possible to proceed with the analysis of the diffraction of TMV gels. Using the theory of Cochran, Crick, and Vand, J. D. Watson [5] was able to show that the structure of TMV was helical with 3n + 1 subunits every 6.9 nm. Watson's best value for n was 10. Independently of each other, R. E. Franklin [6] and D. L. D. Caspar [7] took up the study of the structure of the orientated gels. Franklin was soon to be joined by A. Klug. After Franklin's untimely death in 1958 Klug took over the leadership of the group which was at that time located in Birkbeck College, London, under the aegis of Bernal. In 1962, the virus group moved to the Medical Research Council Laboratory for Molecular Biology in Cambridge, where a great deal of the work described in the following paper was carried out. It is noteworthy that, through all the present-day virus crystallography, one can trace the visionary influence of Bernal and Frankuchen.

Before starting work on TMV, Rosalind Franklin had played a major role in the elucidation of the structure of DNA. One of her motivations in turning to TMV was an interest in the companion nucleic acid RNA. Of the systems containing RNA in an organized form TMV seemed to be ripest for attack. Franklin started out with the ambitious resolve to determine the structure of TMV by means of the then newly discovered method of isomorphous replacement. It was quickly established that the virus had 49 subunits in three turns [8], that the virus was hollow, and that the nucleic acid was located at a radius of 4.0 nm [6]. These results were obtained by analysis of the low-resolution part of the zero layer-line ($<1.0\ nm^{-1}$), where it can be shown that the data are effectively the diffraction pattern from a centrosymmetric structure (ie, the phase problem reduces to a sign problem).

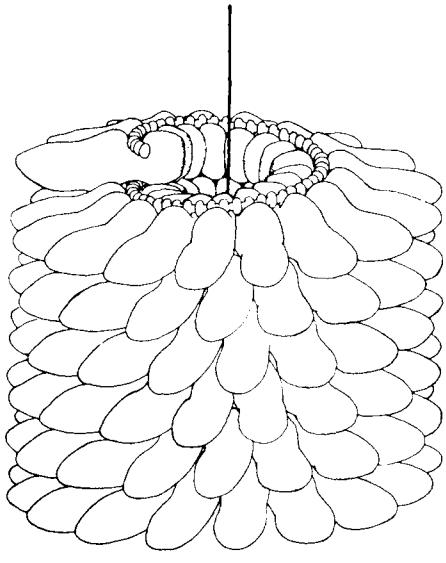

Fig. 1. Diagram of tobacco mosaic virus drawn by D. L. D. Caspar. The six-turn segment shown corresponds to 1/20 the length of the virus. The RNA chain, with three bases per protein subunit, is coiled between the turns of the 2.3-nm pitch helix of protein subunits. The diameter of the RNA is 8.0 nm. There are 16 1/3 subunits per turn. The inner diameter of the particle is 4.0 nm and the outer diameter is 18.0 nm. From Caspar [2], with permission.

The signs of the peaks on the inner part of the zero layer-line were determined independently by Caspar [7] and by Franklin and Holmes [8]. Together with Klug, Franklin showed that the virus surface is marked by a deep helical groove [9]. Using two heavy-atom derivatives, methyl-mercury [10] and osmic acid, she proceeded to analyze the third layer-line in order to arrive at a low-resolution helical projection of the virus. With the death of Rosalind Franklin in 1958, the project lost some impetus. Moreover, in 1958 the necessary methodology, particularly in computer development, was not available. As it turned out, determining the structure of TMV entailed developing new methods in theoretical and practical crystallography and in biochemistry so that nearly 20 years were to elapse before detailed structural information about the RNA and its environment could be obtained. Klug and Holmes continued the structural analysis of the orientated gels by means of the method of isomorphous replacement. In particular, in an unpublished study Klug was able to show that the RNA was single-stranded at a time when it was widely supposed, on the basis of end-group analyses, that the RNA was in many segments. With the establishment of Klug's group in Cambridge the necessary technical milieu became available to extend Franklin's early work.

Microcrystals of protein disks were first reported by Macleod et al [11]. Reproducible crystals were produced by Finch et al [12] in the MRC laboratory, Cambridge. Furthermore, they were able to demonstrate that the symmetry of the disk was 17-fold. This work opened the way to obtaining an atomic model of TMV by means of X-ray crystallography. The disk consists of two rings of protein subunits, each with the same polarity and each containing 17 subunits. The disks crystallize in the orthorhombic space group $P22_12_1$. The molecular weight of the asymmetric unit is $17 \times 2 \times 17{,}400 = 592{,}280$ daltons. The solution of a structure of this size presents formidable problems. To solve the structure of the disk to 0.28 nm resolution it was necessary for Bloomer et al [13] to measure the intensities of 2×10^6 reflections. The analysis at 0.5 nm [14] showed the entire polypeptide chain except for a section within 4.0-nm radius which was not visible. This observation reflects the fact that in the disk (but not in the virus) there is a flexible segment of a length of about 25 residues. The high-resolution structure [13] allowed all the residues to be fitted into the map except for the flexible segments 89–113 and the C-terminal residues 155–158.

The preparation of heavy-atom derivatives for the solution of the phase problem presented major difficulties so that much of the analysis was carried out with a single mercury derivative [13, 14]. Using the formulation of Crowther, Jack [15] was able to demonstrate the power of the noncrystallographic 17-fold symmetry as a supplement to the heavy-atom method for the solution of the phase problem. By setting up the equivalent formulation in real space Bricogne [16] made it possible to use this method for the high-resolution analysis of the structure of the disk [13].

Parallel to this work the analysis of the X-ray fiber diffraction progressed. Here it soon became clear that the limiting factor would be the availability of well-defined single heavy-atom derivatives of the virus. Both genetic and chemical modifications to the virus were used [17–19]. Initially, single heavy-atom derivatives were favored because the location of the heavy atoms in three dimensions from fiber diffraction data proved difficult and only the simplest cases were soluble. In 1968 Holmes moved to Heidelberg and work on the intact virus was transferred to Heidelberg. The determination of the structure of TMV from the X-ray fiber diagrams by means of the method of isomorphous replacement is a unique problem. The data consist essentially of the cylindrical average of the square of the Fourier transform of a single particle [20]. Thus the data at each point in reciprocal space consist of the sum of squares of a number of Bessel function terms, the actual num-

ber of Bessel functions which contribute being determined by the helical symmetry of the particle and the reciprocal space radius of the point in question. At low radii (small scattering angles) only one Bessel function contributes to the intensity. Theoretical studies by Klug et al [21] had shown that in principle it is possible to regenerate the three-dimension structure if the problem of "separating the Bessel functions" can be solved. Franklin and Klug realized that, to a resolution of about 1.0 nm on account of the high symmetry of TMV (49-fold), only one Bessel function could contribute to the intensity. As a consequence it can be shown that to this resolution the intensities are cylindrically symmetrical so that no information is lost by cylindrical averaging. Using this idea Barrett et al [22] were able to calculate a three-dimensional electron density map of the virus with a nominal resolution of 1.0 nm. This map showed the general appearance of the nucleic acid running circumferentially at 4.0 nm radius, and the overall appearance of the subunit including strong radial features which were later shown to be alpha-helices. The structure of the virus extends proximally to a radius of 2.2 nm, in sharp contrast to the situation in the disk. Holmes [23] had shown that it is possible to separate Bessel functions by means of heavy-atom derivatives. The problem of separating Bessel functions by means of isomorphous replacement is a multidimensional analog of the phase problem and can formally be solved by a multidimensional Harker construction. Locating the best intersection of the Harker hyperspheres by a multidimensional search for every point in reciprocal space is a very time-con-

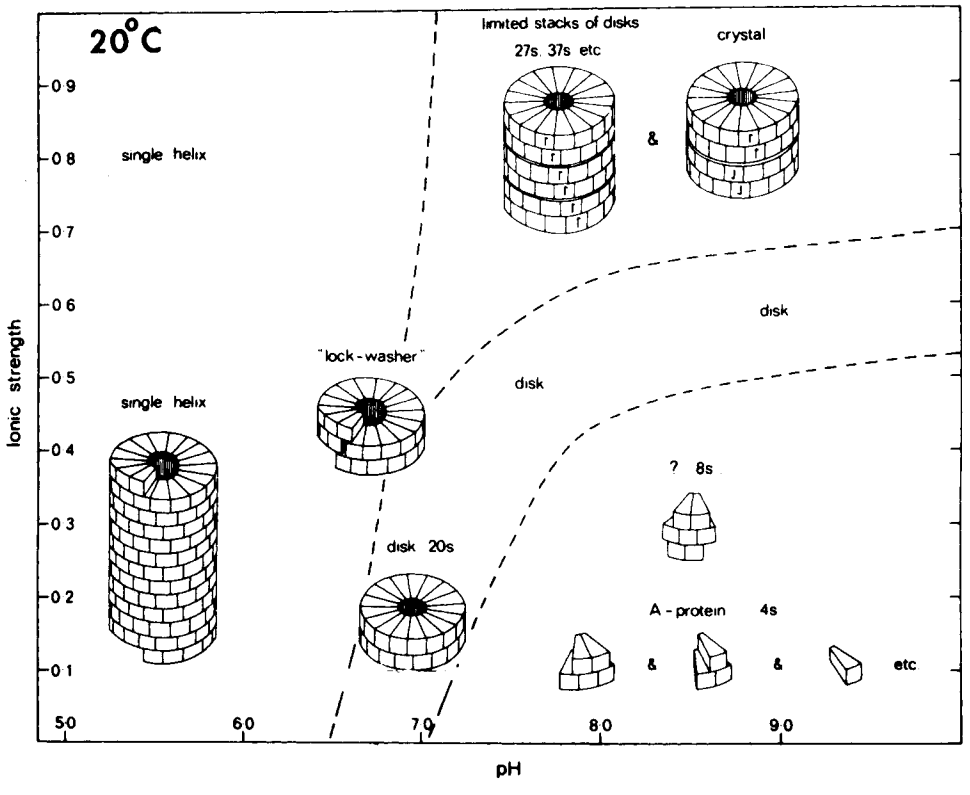

Fig. 2. Diagram showing the ranges of stability of various polymeric species of TMV protein. This is not a conventional phase diagram, since the polymorphic forms shown are in equilibrium. The boundaries indicate regions in which a particular form predominates. The boundaries are approximately correct for a protein concentration of 5 mg/ml at 20°C. From Klug and Durham [29], with permission.

suming computing problem. An algebraic solution was developed by Stubbs and Diamond [24]. Using this method Holmes et al [25] analyzed the data out to a resolution of 0.67 nm, which is the two-Bessel-function limit. The resulting map showed some of the polypeptide chain, in particular the two radial alpha-helices which are such a prominent feature of the structure of the TMV subunit. It allowed a preliminary chain tracing (the 0.5 nm resolution map of the protein disk showed subsequently that the alpha-helices had been correctly identified but not correctly linked together). Further work admitting three Bessel functions, by Stubbs, Warren, and Holmes [26], yielded an electron density map at 0.4 nm resolution by the use of six heavy-atom derivatives, which taken together with the detailed map of the subunit in the disk produced by Bloomer et al [13] yields a considerable amount of stereochemical information about the nature of the protein-RNA interaction in TMV assembly.

POLYMORPHISM OF TMV COAT PROTEIN

The isolated nucleoprotein particle can readily be taken apart [27] and reassembled [28]. In addition, a number of polymorphs of the coat protein have been described [29] (Fig. 2). Moreover, Butler and Klug [30] were able to show that the disk is the main precursor for the initiation and growth of the virus.

The major polymorphic forms of the protein *without* nucleic acid are:

the A-protein (predominantly a trimer);
the disk (containing two rings of 17 subunits);
a rod-like structure made from stacked disks;
a helical virus-like structure (16 1/3 subunits per turn);
a helical virus-like structure (17 1/3 subunits per turn [31]).

At neutral pH and low ionic strength the disk is the majority species. By raising the salt concentration the formation of the stacked disk is encouraged. The stacked disk may apparently be trapped by mild proteolytic cleavage of the flexible segment in a form which can no longer be taken apart [32]. This demonstrates the important role of the flexible segment in the aggregation and disaggregation of the disk. On lowering of the pH, the disks build a virus-like helix even in the absence of nucleic acid. This is thought to come about through the formation of an intermediate "lockwasher" [29]. The lockwasher may form by dislocating along either of two intersubunit boundaries to form a 16 1/3 or 17 1/3 helix [31]. In fact both forms are found, which lends credence to the idea that the helix can be built from dislocated disks. In the presence of RNA only the 16 1/3 form is found.

A remarkable feature of all these polymorphs is the conservative nature of the side-to-side bonding between the subunits, which scarcely alters. In contrast, the up-and-down bonding between subunits is markedly different between the disk forms and the helical forms [14] (Fig. 3). The subunits are 0.3 nm further apart in the disk than in the helix and are displaced 1.0 nm circumferentially.

ASSEMBLY OF THE VIRUS

During the assembly of the virus the coat protein has to specifically recognize the viral nucleic acid and at the same time has to be able to tolerate the variation of sequence found when encapsulating the RNA. Butler and Klug [30] showed by kinetic analysis that

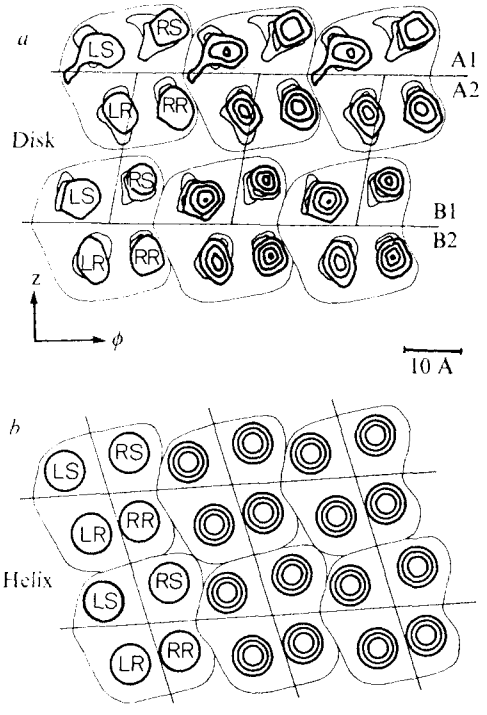

Fig. 3. Packing of subunits and alpha-helical rods in the protein disk and in the virus. a, Superposed cylindrical sections of the electron density map of the disk at radii 5.7, 5.85, and 6.0 nm. The outline of the subunit is shown. b, Schematic diagram of the arrangement of subunits in the helix at the same radius (from Champness et al [14], with permission). Note the change in the surface lattice in going from the disk to the helix, which involves a movement of about 1.0 nm circumferentially at this radius. The side-to-side contacts remain the same.

the disk is the major species responsible for coating the RNA during in vitro assembly. Assembly starts by the insertion of a specific RNA loop (the initiation loop) into the disk [33]. A specific sequence of the RNA (the assembly origin) binds to the disk [34–36], thereby mediating the formation of the lockwasher and initiating the growth of the helix. The assembly origin of the RNA is about 100 nucleotides long and the initial binding is thought to occur between the turns of the disk. Close to the center of the assembly origin is the sequence AGAAGAAGUUGUUGAUGA. In this and the surrounding sequences there is a strong tendency for every third residue to be G. The assembly origin has a very low C content and occurs about 15% of the length away from the 3′ end. Growth takes place fastest in the 5′ direction [37, 38]. Growth in the 3′ direction goes on at the same time but rather more slowly and possibly by a different mechanism [38]. Rather unexpectedly, as growth proceeds towards the 5′-hydroxyl end the 5′ end is dragged through the central hole of the nascent virus (Fig. 4) so that both the 3′ and 5′ ends are initially at the same end of the virus [39, 40].

STRUCTURE OF THE DISK

The general appearance of the subunit, as deduced from the high-resolution studies of Bloomer et al [13], is shown in Figure 5. Both the C and N termini are located distally.

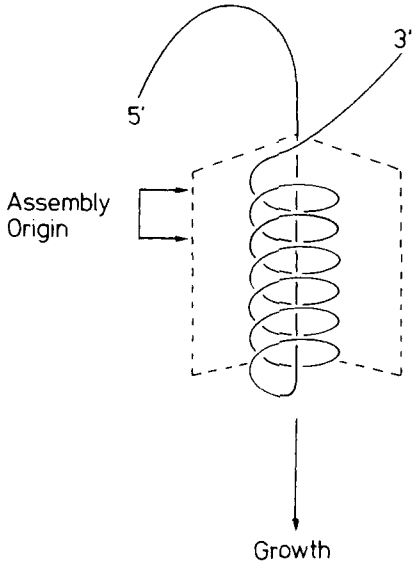

Fig. 4. Diagram of the encapsulation of RNA during the growth of TMV particles. The dominant direction of growth (indicated by an arrow) is downwards from the assembly origin (towards the 5' end). The RNA is pulled through the hollow core of the nascent virus. From Butler et al [39], with permission.

Starting from the N terminus the chain builds a short helical structure between residues 8 and 15. Then follows a short section of extended chain (16–18) which forms part of the distal beta-pleated sheet. The chain continues into the LS (left-slew) helix (20–32) which runs proximally to a radius of 5.0 nm. After a tight turn (33–38) the chain proceeds distally along the RS (right-slew) helix (38–48), which is parallel to the LS helix. The chain now forms part of a beta-sheet. At 74 the RR (right-radial) helix starts. This helix, which is tilted about 20° to the horizontal, runs proximally and up to a radius of about 4.5 nm, where the density fades out. Then follows a flexible segment containing 24 residues, which has been shown by NMR investigations [41] to be highly mobile. After the flexible segment (89–113) the LR (left-radial) helix runs distally roughly horizontally from a radius of 4.0 nm to 7.0 nm (114–134). Then the chain runs through a short alpha-helical segment to the C terminus. The last four residues (155–158) are not visible and are in a random coil. The four main alpha-helices include about 60 residues of the total of 158. The helices are packed as in a segment of a left-handed, four-stranded supercoiled bundle with hydrophobic contacts along the center. The distal ends of the four helices are connected transversely by a strip of beta-sheet shown schematically in Figure 6. Residues 53–54, 16, 18, 68–71, 137–139, and 135–136 are involved. In the central region of each subunit on the distal side of the beta-sheet is a cluster of aromatic residues, which gives rise to a continuous belt of hydrophobic interactions encircling each ring of the disk.

STRUCTURE OF THE VIRUS

The structure of the subunit in the virus is substantially the same as the disk except for the flexible loop and the vertical intersubunit contacts. Proximally from 4.0 nm radius, where the density in the disk fades out, one sees the LR and RR helices continuing to a

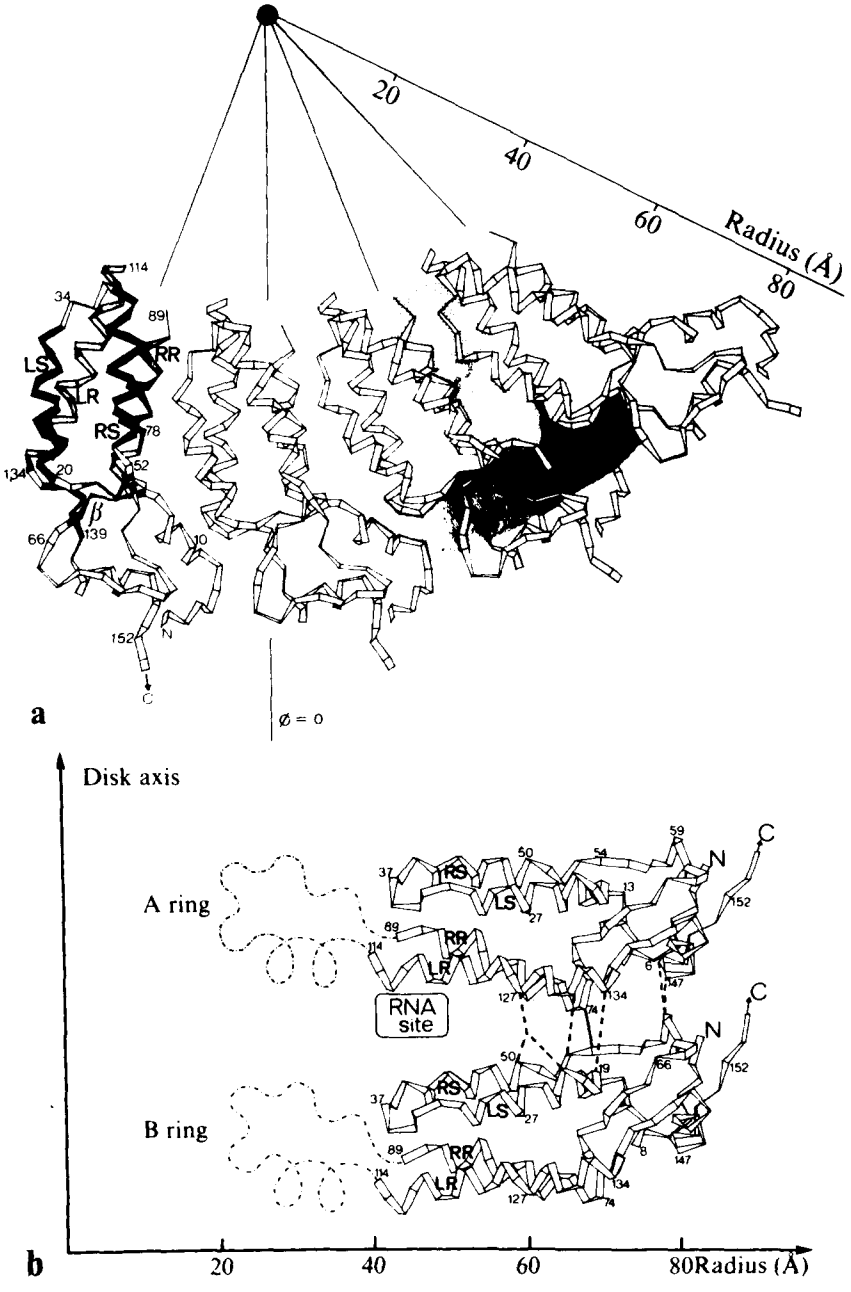

Fig. 5. Plan (a) and elevation (b) of the TMV disk. a, Four adjacent subunits of the A ring are viewed from above. The N and C termini of the chain are marked on the extreme left subunit together with the four main helices, left and right slewed (LS and RS) in the upper half of the molecule with right and left radial (RR and LR) below, and the beta-sheet, which connects all four helices at their distal ends. The subunit interface contains alternating patches of polar residues (shown by a stippling) and hydrophobic residues (shown with solid shading). Distally from the beta-sheet is the hydrophobic girdle, which extends circumferentially across the whole width of the subunit. From Bloomer et al [13], with permission. b, Side view through a sector of the disk showing the disposition of the subunits in the A and B rings and the axial contacts between them. There are three regions of contact, indicated by a solid line for the hydrophobic contact of pro 54 with ala 74 and val 75, by dashed lines for the hydrogen bonds between thr 59 and ser 147 and 148, and further dashed lines for the extended salt-bridge system. All these bonds are broken in the transition to the helical structure of the virus. The proximal region of the subunit, which has no structure in the disk form, is shown schematically by a broken line in approximately the configuration it takes in the virus. From Bloomer et al, with permission.

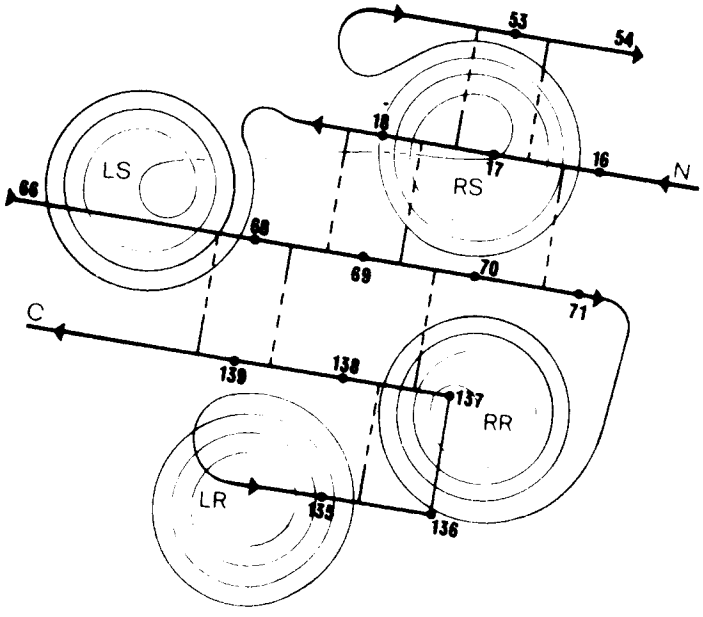

Fig. 6. Schematic view of the antiparallel beta-sheet. Shown is the connectivity of the four strands and of the four principle alpha-helices. The direction of view is radially towards the center. The beta-sheet is a narrow, twisted strip which extends across the subunit, bracing together the four alpha-helices. At the bottom of the beta-sheet is a type II beta-bend involving the invariant gly 137. From Bloomer et al [13], with permission.

low radius (2.5 nm) (Fig. 7), where they are joined together by a strong vertical column of density (the V-column) which was identified by Stubbs et al [26] as an alpha-helix, although more recent studies show this identification to be somewhat uncertain. In particular, the number of residues involved is less than was proposed by Stubbs et al. In addition to this density, which can unequivocally be ascribed to the protein, there is a ribbon of density running roughly circumferentially at a radius of 4.0 nm *underneath* the LR helix which can be ascribed to the nucleic acid. Recent studies by Mandelkow et al [42] of the structure of the helical protein aggregate *without* RNA substantiate the assignment of this density to the RNA. The studies of Mandelkow et al have led to some small revisions of the RNA conformation from the form reported by Stubbs et al, but the essential features remain unchanged. A diagram of the RNA conformation is shown in Figure 8. Each protein subunit binds three bases which are numbered 1, 2, and 3 from the 5′ end. Two of the bases appear to be in the *anti* conformation and one in the *syn*, but this is not very well determined. The sugar puckers are all C3′ endo. The direction of the nucleic acid agrees with that independently determined by Wilson et al [43]. The bases form a claw-like structure around the LR helix (Fig. 9). Two bases are at a radius of about 3.8 nm and one base is at a larger radius (4.5 nm). The phosphates are clustered around a radius of 4.0 nm and appear to bond to the adjoining RS and RR helices from the neighboring subunit in the next turn down. Thus the RNA binding site lies between the rings of a disk, in agreement with the model for assembly proposed by Butler et al [33].

RELATIONSHIP BETWEEN THE DISK AND VIRUS MAPS

Gilbert [44] estimated the shift involved in changing from the disk to the helix by assuming that the protein subunit packing at a radius of 8.0 nm stayed the same in the azimuthal direction and that no rearrangement of the subunit structure took place. One ob-

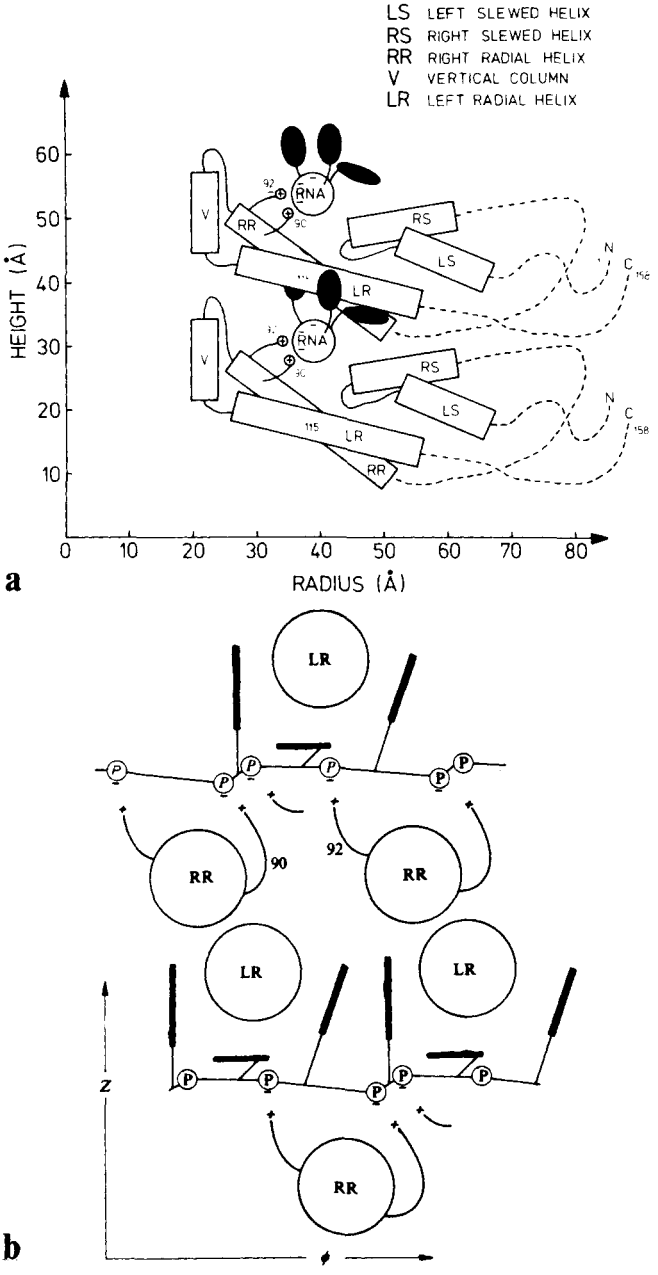

Fig. 7. Schematic view of the inner part of the subunit structure in the virus derived from the electron density maps at 0.4-nm resolution calculated from X-ray fiber diffraction patterns of the virus. a, A sector of the virus viewed circumferentially showing the extended LR and RR helices and the V-column joining them together. This part of the structure is not seen in the disk, although it is represented by high density in the helical virus. Comparison with the disk structure shows that the LR helix should be displaced distally by one turn so that the bottom of the V-column will probably occur at about residue 108 rather than 112. An important component of the RNA binding site is the interaction with the LR helix between residues 114 and 123. This part of the LR helix already exists in the disk. The detailed structure of the V-column is at present unclear. b, Diagram of the RNA binding site at a radius of 4.0 nm. The direction of view is radially towards the center. Note how the bases (thick lines) group around the LR helix in the shape of a claw. The base lying horizontally is at a larger radius (circa 4.8 nm). The phosphate groups form salt bridges with the subunit underneath so that the RNA binding site lies between turns of the virus helix. a and b from Stubbs et al [26], with permission.

Fig. 8. The RNA configuration in TMV. Two full repeats (six nucleotides) are shown. The 5′ end is to the left. The direction of view is from above with the center of the virus at the top of the diagram. The mean radius of the RNA (4.0 nm) is indicated with a broken-line circle. The nucleotides have been numbered 1, 2, and 3. The phosphate groups are shown with a heavy line. Bases 1 and 3 rise up vertically out of the plane of the diagram and have been omitted for clarity. The electron density associated with the bases is shown by broken-line contours. From Mandelkow et al [42].

tains the value 0.32 nm for the radial displacement necessitated by changing the symmetry from 17-fold to 16 1/3-fold. Comparing heavy-atom positions in the virus and in the disk supports the assumption that the change from disk to helix is approximately a rigid body transformation [45] so that Gilbert's assumptions seem to be justified.

Bloomer [13] has analyzed the relationship between the A-disk and the B-disk and the virus and has shown that, in addition to the radial movement, the subunits must be rotated by 10° (A-disk) or 20° (B-disk) about an axis perpendicular to the common disk or helix axis in order to bring them into the same orientation as the virus. Recently Bricogne and Bloomer, in an unpublished study, have determined the transformation necessary to bring the A- and B-disk density into coincidence with the virus density by a least-squares procedure.

The tilt of the subunits in the virus helix gives rise to the characteristic "Christmas tree" appearance of the virus particle [43], which may be used to define the polarity of the virus. The point of the Christmas tree is taken as "up" in structural studies and "down" in assembly and elongation studies!

THE RNA BINDING SITE

The following account of the binding site should be taken as plausible but not proven.

If we apply Bricogne and Bloomer's transformation to the LR helix we find that it amounts to a translation of about 0.32 nm in the radial direction in the neighborhood of the nucleic acid. Applying this transformation to the LR helix in the A-ring of the disk results in the relationship between the LR helix and RNA shown in Figure 10. Comparing the resulting structure with the virus electron density indicates that only minor revisions will be necessary to take account of local distortions. Comparing the model so produced with the model fitted to the helix electron density by Stubbs et al (Fig. 7), we

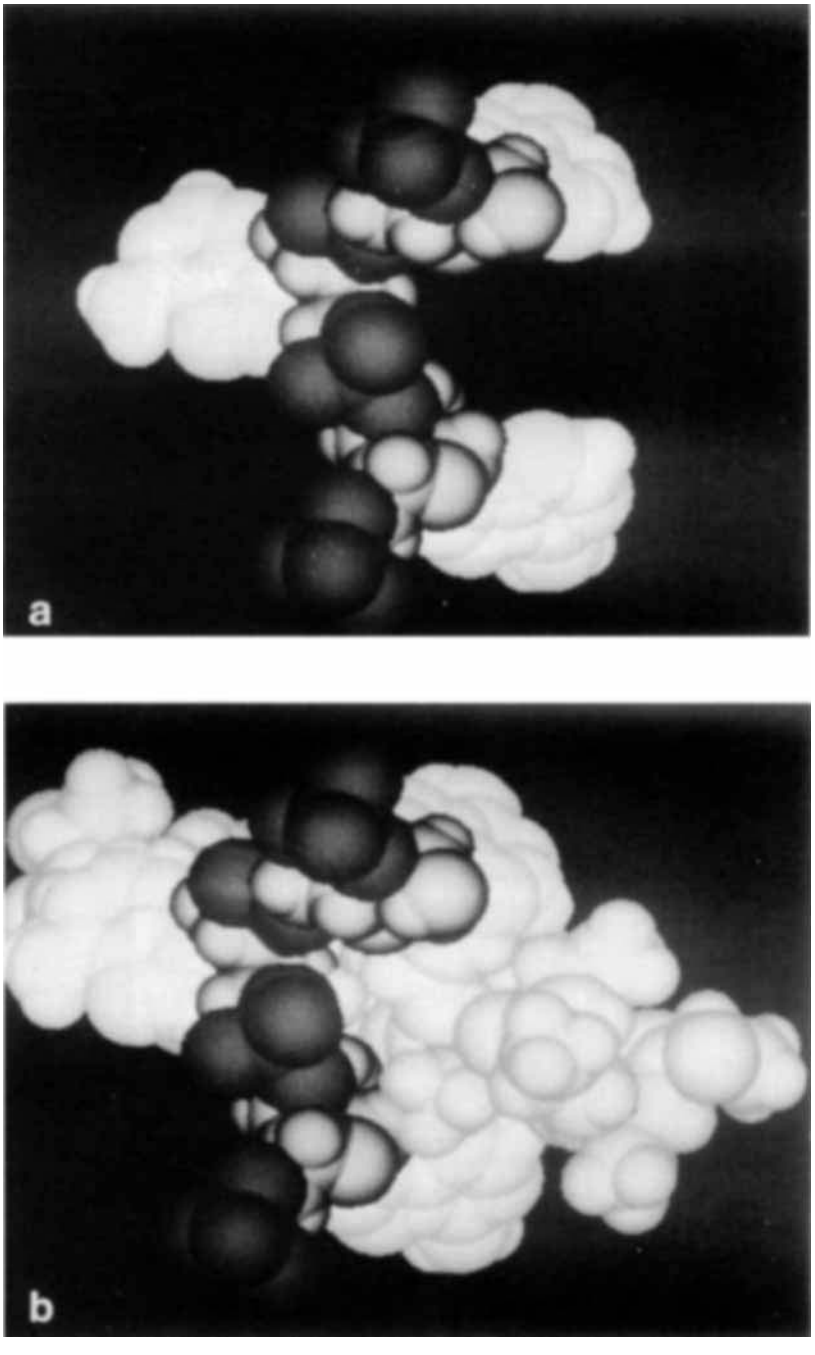

Fig. 9. A computer drawing (Dr. R. J. Feldmann, NIH from coordinates provided by Dr. G. Stubbs, Brandeis) of the relationship between the LR helix and the RNA claw. a, The three-nucleotide repeat in the claw configuration. The bases are white, the sugars light grey, and the phosphate groups dark grey. The direction of viewing is from underneath with the center of the virus on the right. b, The RNA claw gripping the LR helix.

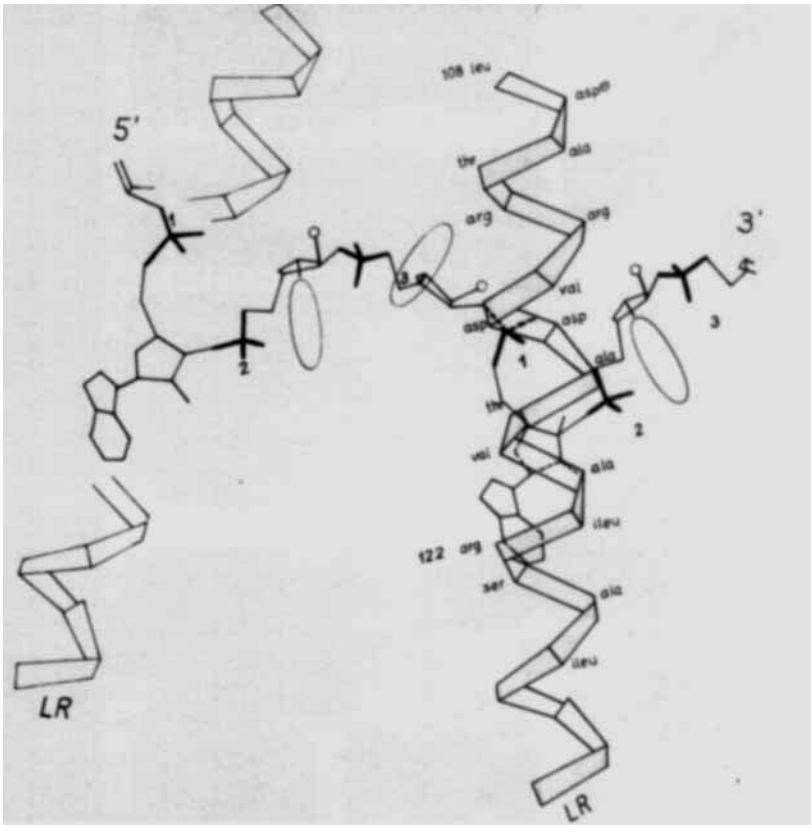

Fig. 10. The RNA binding site. Data are from Bloomer et al [13] and Mandelkow et al [42]. Shown are the RNA and the LR helix. The direction of view is from above with the center of the virus at the top of the diagram (as in Fig. 8). Two repeats are shown. The LR helix runs above the RNA. In the left repeat, part of the LR helix has been omitted for clarity. The bases which rise up out of the plane of the diagram are shown by ellipses. The 2′ OH groups are shown with small circles. Note how the 2′ OH groups of bases 1 and 3 come close to Asp 115 and 116 of the LR helix.

find that the transformed disk structure is one turn further out (3.6 residues) than the Stubbs, Warren, and Holmes model. A similar shift has already been proposed by Bloomer et al [13]. Given the high quality of the electron density map of the disk, this comparison strongly suggests that the virus model needs revision, namely, that the LR helix in the virus model should be translated one turn distally.

With the virus model adjusted in this way, some interesting features of the RNA binding site reveal themselves: An important component of the binding site seems to be provided by the invariant pair of aspartate residues (115, 116). These probably hydrogen-bond to the RNA ribose 2′ OH on residues 1 and 3. Such a ribose binding site is typical of alcohol dehydrogenase, lactate dehydrogenase, and malate dehydrogenase [46–48]; here this binding motif may be used twice per protein subunit. The accompanying bases find themselves in hydrophobic pockets close up against the alpha-helix, as was suggested by Stubbs et al [26]. For base 1 the methylene chain of Arg 113 appears to be an important component of the hydrophobic site. The other component of the site is the β-carbon of Asp 115. The

guanidinium group of Arg 113 might also be involved in a salt bridge to a phosphate group. Base 3 appears to bind against Val 114 and Ala 117. Base 2 is at a larger radius and may form a hydrogen bond to Ser 123. The sugar of residue 2 seems to lie face down against a hydrophobic surface formed by residues 119 (Val) and 120 (Ala). If the 2′ OH is able to form a hydrogen bond, it must be to the neighboring slewed hairpin joining the LS or RS helices from the subunit underneath, possibly from Asn 33.

The effect of the tight binding of residues 1 and 3 to the LR helix is to pucker the RNA so that the phosphate residues are grouped together, thereby making them a target for the formation of salt bridges with arginines. Stubbs et al [26] have suggested that the salt bridges are formed with arginine residues 90 and 92 from the RR helix of the subunit directly underneath, and possibly with residue 41 from the RS helix of this subunit. The assignment of the invariant residues 90 and 92 to the phosphate binding site is plausible, but higher resolution studies of the virus structure are needed before such an assignment can be proven. Unfortunately, this part of the RR helix cannot be seen in the disk structure [13]. The disk structure indicates that Arg 41 is not such a strong candidate for the phosphate binding site since it lies on the wrong face of the RS helix.

A HYPOTHETICAL TRIGGER MECHANISM FOR THE DISK-HELIX TRANSITION

The surprising result from the studies summarized above is that the major component of the RNA binding site in TMV is one alpha-helix, the LR helix, which binds three bases. In the disk structure, which is the template for binding RNA, this helix continues as a finger proximally to a radius of about 4.0 nm (residue 114) before the electron density peters out. This finger could nucleate the binding of the RNA by means of three kinds of interaction: 1) a stereospecific interaction of the aspartate groups 115, 116 with two of the ribose groups; 2) a hydrophobic interaction with the three bases forcing them into the shape of a claw around the alpha-helix (Fig. 9); 3) specific hydrogen bonding which is still to be determined. Present studies suggest a possible hydrogen bond between ser 123 and base 2 which would favor a hydrogen bond acceptor in this position.

The binding is accompanied by the formation of salt bridges between Arg 90 and 92 from the neighboring subunit one turn deeper in the virus helix and the phosphate groups of the nucleic acid. Initially 90 and 92 are part of the random-coil segment adjoining the RR helix, but apparently through the binding to the phosphate groups this structure is stabilized and as a result the V-column and the rest of the LR and RR helices, which are not seen in the disk, can build. However, the V-columns from adjoining subunits are in close contact and it seems probable that they can only fit together in one way, namely, as in the virus helix. The change in relative heights in passing from the disk to the helix is 0.14 nm per subunit, and it is conceivable that the packing of the V-columns could not tolerate this much distortion. As a result, therefore, of the packing of the V-columns the disk structure may be strained and may attempt to transform into the helix. The contact between the A- and B-rings of the disk is mediated by a complex system of hydrogen bonds (Fig. 6) which must break when the helix is formed. The disk-helix transition is therefore reminiscent of the R-T transformation in hemoglobin [49]. The disk is the low-affinity form for RNA because one cannot make the salt bridges to the phosphates without allowing the V-column to build, thereby destabilizing the disk. The helix is the tight-binding form. However, to reach the helix it is necessary to break the network of hydrogen bonds between the A and B rings of the disk. The balance between these effects determinés the binding of the RNA and makes the RNA binding highly cooperative. Perhaps it is relevant in this context that

Lomonossoff and Butler [50] report that the incorporation of RNA during in vitro reassembly is quantized, the quantum of binding being close to 50 bases or 100 bases. This is essentially the number of bases in one or two turns, which would seem to show that binding proceeds through the addition of single or double rings, thereby supporting the idea that the disk is the major precursor of virus growth. Through such cooperativity one could also explain the specificity of the binding. Two properties of the assembly origin seem noteworthy: the requirement that every third residue should be G; and the requirement that the sequence should be long. If we postulate a binding mechanism whereby the selectivity for any one site is small (eg, AGU may have twice the binding constant of UUU) but where the *simultaneous* binding to 16 sites produces the disk-helix transition, then one sees at once why the assembly origin can be so specifically recognized – it has the selectivity 2^{16}.

Stubbs et al [26] have pointed out that the inner structure of the virus (the V-column and the adjoining pieces of LR and RR helix which build on binding RNA) is destabilized by a high concentration of carboxyl groups in this region, which would probably explain the anomalous pKs appearing during the disk-helix transition [2, 49]. The important function of the energetically unfavorable carboxylate-carboxylate interactions may be to help keep part of the RNA binding site in a random-coil configuration until the RNA binds.

ACKNOWLEDGMENTS

I gratefully acknowledge the help I have received from Dr. Gerald Stubbs and Dr. Anne Bloomer, in particular their generously making unpublished results available to me. I am grateful to Drs. Stubbs, Bloomer, and Klug for reading the manuscript and for making many helpful suggestions.

REFERENCES

1. Anderer FA: Adv Protein Chem 18:1, 1963.
2. Caspar DLD: Adv Protein Chem 18:37, 1963.
3. Bernal JD, Fankuchen I: J Gen Physiol 25:111, 1941.
4. Cochran W, Crick FHC, Vand V: Acta Crystallogr 5:581, 1952.
5. Watson JD: Biochim Biophys Acta 13:10, 1954.
6. Franklin REF: Nature (Lond) 177:929, 1956.
7. Caspar DLD: Nature (Lond) 177:928, 1956.
8. Franklin REF, Holmes KC: Acta Crystallogr 11:213, 1958.
9. Franklin REF, Klug A: Biochim Biophys Acta 19:403, 1956.
10. Fraenkel-Conrat H: In Benesch R, Benesch RE (eds): "Sulphur in Proteins." New York: Academic, 1959, p 339.
11. Macleod R, Hills GJ, Markham R: Nature (Lond) 200:932, 1963.
12. Finch JT, Leberman R, Chang Y-S, Klug A: Nature (Lond) 212:349, 1966.
13. Bloomer AC, Champness JN, Bricogne G, Staden R, Klug A: Nature (Lond) 276:362, 1978.
14. Champness JN, Bloomer AC, Bricogne G, Butler PJG, Klug A: Nature (Lond) 259:20, 1976.
15. Jack A: Acta Crystallogr A29:545, 1973.
16. Bricogne G: Acta Crystallogr A32:832, 1976.
17. Wittmann HG: Z Vererbungslehre 95:333, 1964.
18. Perham RN, Thomas JO: J Mol Biol 62:415, 1971.
19. Gallwitz U, King L, Perham RN: J Mol Biol 87:257, 1974.
20. Franklin RE, Klug A: Acta Crystallogr 8:777, 1955.
21. Klug A, Crick FHC, Wyckoff HW: Acta Crystallogr 11:199, 1958.
22. Barrett AN, Barrington Leigh J, Holmes KC, Leberman R, Mandelkow E, von Sengbusch P: Cold Spring Harbor Symp Quant Biol 36:433, 1971.

23. Holmes KC: PhD Thesis, University of London 1959.
24. Stubbs GJ, Diamond R: Acta Crystallogr A31:709, 1975.
25. Holmes KC, Stubbs GJ, Mandelkow E, Gallwitz U: Nature (Lond) 254:192, 1975.
26. Stubbs GJ, Warren S, Holmes KC: Nature (Lond) 267:216, 1977.
27. Schramm G: Z Naturforsch 2b:112,249, 1947.
28. Fraenkel-Conrat H, Williams RC: Proc Natl Acad Sci USA 41:690, 1955.
29. Klug A, Durham ACH: Cold Spring Harbor Symp Quant Biol 36:449, 1971.
30. Butler BJG, Klug A: Nature New Biol 229:47, 1971.
31. Mandelkow E, Holmes KC, Gallwitz U: J Mol Biol 102:265, 1976.
32. Durham ACH FEBS Lett 25:147, 1972.
33. Butler PJG, Bloomer AC, Briconge G, Champness JN, Graham J, Guilley H, Klug A, Zimmern D: In Markham R, Horne R (eds): "Structure-Function Relationships of Proteins." 3rd John Innes Symposium. Amsterdam: North Holland-Elsevier, 1976.
34. Zimmern D, Butler PJG: Cell 11:455, 1977.
35. Zimmern D: Cell 11:463, 1977.
36. Jonard G, Richards KE, Guilley H, Hirth L: Cell 11:483, 1977.
37. Zimmern D, Wilson TMA: FEBS Lett 71:294, 1976.
38. Lomonossoff P, Butler PJG: Eur J Biochem 93:157, 1979.
39. Butler PJG, Finch JT, Zimmern D: Nature (Lond) 265:217, 1977.
40. Lebeurier G, Nicolaieff A, Richards KE: Proc Natl Acad Sci USA 74:149, 1977.
41. Jardetzsky O, Akasaka K, Vogel D, Morris S, Holmes KC: Nature (Lond) 273:564, 1978.
42. Mandelkow E, Stubbs GJ, Warren S: J Mol Biol (Submitted for publication).
43. Wilson TMA, Perham RN, Finch JT, Butler PJG: FEBS Lett 64:285, 1976.
44. Gilbert PFC: PhD thesis, University of Cambridge, 1970.
45. Graham J, Butler PJG: Eur J Biochem 83:528, 1978.
46. Branden CI, Eklund H, Nordstrom B, Boiwe T, Soderlund G, Zeppezauer E, Ohlsson F, Akeson A: Proc Natl Acad Sci USA 70:2439, 1973.
47. Smiley IE, Koekack R, Adams MJ, Rossmann MG: J Mol Biol 55:467, 1971.
48. Hill E, Tsernoglou D, Webb L, Banaszak LJ: J Mol Biol 72:577, 1972.
49. Durham ACH, Klug A: Nature New Biol 229:42, 1971.
50. Lomonossoff GP, Butler PJG: J Mol Biol 126:877, 1978.