Journal of Supramolecular Structure 12:177–183 (1979)
Biological Recognition and Assembly 141–147

Structural Comparisons of the Aggregates of Tobacco Mosaic Virus Protein

Gerald Stubbs, Stephen Warren, and Eckhard Mandelkow

Rosenstiel Center, Brandeis University, Waltham, Massachusetts 02254 (G.S., S.W.) and Abteilung Biophysik, Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, 69 Heidelberg, Federal Republic of Germany (E.M.)

The coat protein of tobacco mosaic virus forms numerous aggregates, including the small A-protein, the disk, and two helical forms. The structures of the disk, the helical protein forms, and the virus are compared. Most of the differences are in the conformation of the chain between residues 89 and 113, which lies in the region of protein at the center of the virus, inside the RNA. It is disordered in the disk, but has a fixed conformation in the virus and the protein helices. The differences between the virus and the two helical protein forms are largely in the conformations of arginines and carboxylic acids in this region.

Key words: tobacco mosaic virus protein, X-ray diffraction, protein structure

INTRODUCTION

Tobacco mosaic virus (TMV) is a rod-shaped plant virus, 3,000 Å long and 180 Å in diameter. The coat protein subunits (MW 17,500) form a helix of 49 subunits in three turns. A single strand of RNA follows the basic helix at a radius of about 40 Å, with three nucleotides bound to each protein subunit. Early work on structure and assembly of TMV has been reviewed by Caspar [1].

The coat protein of TMV forms a large number of aggregates, which fall into three main groups: A-protein, disk, and helical forms. The interconversions between aggregates have been summarized by Durham and Klug [2]. The A-protein is a mixture of oligomers (3–12 subunits) present at low ionic strength and high pH. Lowering the pH to near neutral or raising the ionic strength produces the disk, which consists of two layers of 17 subunits each [3]. The layers face the same way, but further polymerization is hindered by a pairing interaction between them [1, 4]. At pH below neutral the protein forms long helices [5] with a morphology similar to that of the virus [6]. There are two forms of helical aggregate (as well as the virus itself): The A form has 16-1/3 subunits per turn, like the virus, whereas the B form has 17-1/3 subunits per turn [7]. The existence of two separate forms was found by X-ray diffraction from oriented gels, and the technique for preparing such gels [8] is such that it has not been possible to correlate the conditions of formation with the type of helix formed.

Received April 9, 1979; accepted September 9, 1979.

0091-7419/79/1202-0177$01.70

X-ray structural studies have been made of the intact virus [9], the disk [10], and both forms of protein helix. Some features of the protein helix structures will be described in this paper, but full details will be given elsewhere (Mandelkow, Stubbs, and Warren, in preparation). The virus structure was determined at a resolution of 4 Å by X-ray diffraction from oriented gels. That part of the structure containing the RNA binding site and some of the features which are important in virus assembly are shown in Figure 1. A more complete illustration of the virus structure is given in Stubbs et al [9: Fig. 2a]. The virus contains four approximately radial α-helices, termed [11] the left and right radial (LR and RR) and the left and right slew (LS and RS). LS and RS lie on top of LR and RR, at angles of 10–20°, as in a coiled-coil of α-helices [10]. LR and RR are connected at the inside wall of the protein (the hole down the center of the virus) by a short length of chain called V [9]. V contains two or three turns of rather irregular α-helix. The protein at higher radius than these five helices has not yet been well described for the virus, because the resolution of the map falls to about 5.5 Å at radii between 60 Å and 80 Å. At a radius of 40 Å the RNA is bound to the protein. The three phosphate groups form ion pairs with arginines 90, 92, and either 41 or 113. The bases lie flat against the hydrophobic surface of the LR helix, forming a shape like a saddle. In the region between the RNA and V, there are six carboxylic acids, which have been called the "carboxyl cage" [9]. Four of these (Glu 95, Glu 97, Glu 106, and Asp 109) are near a site which binds divalent cations, and could be a physiologically significant metal-binding site, perhaps binding calcium [9, 12]. The other two (Asp 115 and Asp 116) are about 10 Å away, nearer the RNA. This arrangement is reminiscent of the double metal-binding sites of thermolysin [13] and concanavilin A [14, 15], although the two sites in TMV are further apart.

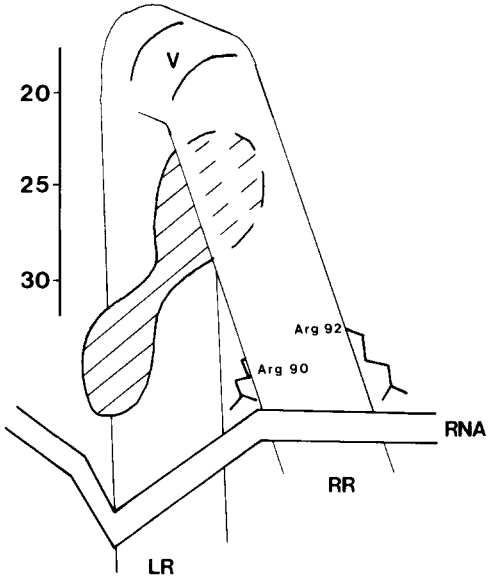

Fig. 1. RNA binding site in TMV (schematic), viewed in the direction parallel to the helix axis. The bar at the side runs in a radial direction, and indicates distance from the axis. The shaded region represents the carboxyl cage. V and parts of LR and RR are shown, together with the backbone of the RNA. Bases of the RNA are not shown, but if they were, the bases of the RNA strand *behind* the one shown would bind to LR.

The disk crystallizes [3], and its molecular structure has been determined to 2.8 Å resolution [10]. Although the intersubunit arrangement is different, the internal structure of the disk subunit [11: Fig. 1] is very similar to that of the virus. The major difference is that the loop of protein containing V, residues 89–113, is disordered in the disk crystal [11] and has been shown to be very flexible in solution [16]. A feature which was not evident in the virus map because of the lower resolution but is presumably common to both structures is the "hydrophobic girdle," a concentration of aromatic and proline residues about 75 Å from the center of the disk [10].

It is evident that the morphology and general fold of the protein chain are well understood for all the aggregates. However, the differences in molecular structure between the aggregates are less clear. This paper addresses these differences and attempts to describe some of them.

METHODS

The methods used to calculate the electron density maps of the protein helices will be described briefly here and more fully elsewhere. Samples of repolymerized TMV protein were prepared as described by Mandelkow et al [7] and made into oriented gels [8]. These gels were of as high a quality as those used to determine the virus structure [9]. Diffraction patterns were recorded on film, measured on a computer-controlled flat-bed scanner and corrected for geometric factors [17]. These diffraction patterns were very similar to those of the virus but fell into two classes, A and B [7]. Electron density maps of both forms were calculated to a resolution of 4 Å using intensities from the protein helices and phases from the virus. The most useful maps were $(2|F_{obs}|-|F_c|)\exp(i\alpha_c)$ Fourier-Bessel syntheses. $|F_{obs}|$ was obtained from the diffraction patterns of the protein helices. $|F_c|$ and α_c were usually obtained from the virus map in the following way: A subunit of TMV in the map was defined by an envelope and transformed to calculate $|F_c|$ and α_c. Various envelopes were used, deleting the RNA or regions of protein which were of particular interest. Deleting such a protein region from the calculated structure factors ensures that the map will not be biased towards identity with the virus. For comparison, one map of the A form was calculated using the measured virus structure factors as $|F_c|$ and the virus isomorphous replacement phases [9] as α_c. This was possible because the A form is isomorphous with the virus.

The protein helices were compared with the virus by careful sheet-by-sheet comparisons of the electron density maps. The disk was compared with the virus using the published descriptions of the disk structure [10, 11].

RESULTS

The principal differences between the various aggregates are shown schematically in Figure 2. Many of the differences between the disk and the virus have been described by Champness et al [11]. The RNA is not part of the disk structure. In its place there is a large open area, formed by the pairing interaction between the layers, which tilts the layers apart at low radius. It has been described as a pair of jaws ready to accomodate the RNA. From the RNA binding site to the center of the disk there are no structural features in the map, so that the hole in the center of the disk appears to be 80 Å across, compared with 40 Å in the virus. Part of the hole must be occupied by a disordered chain from residue 89 to residue 113. By contrast, this chain is fully ordered in the virus: Two

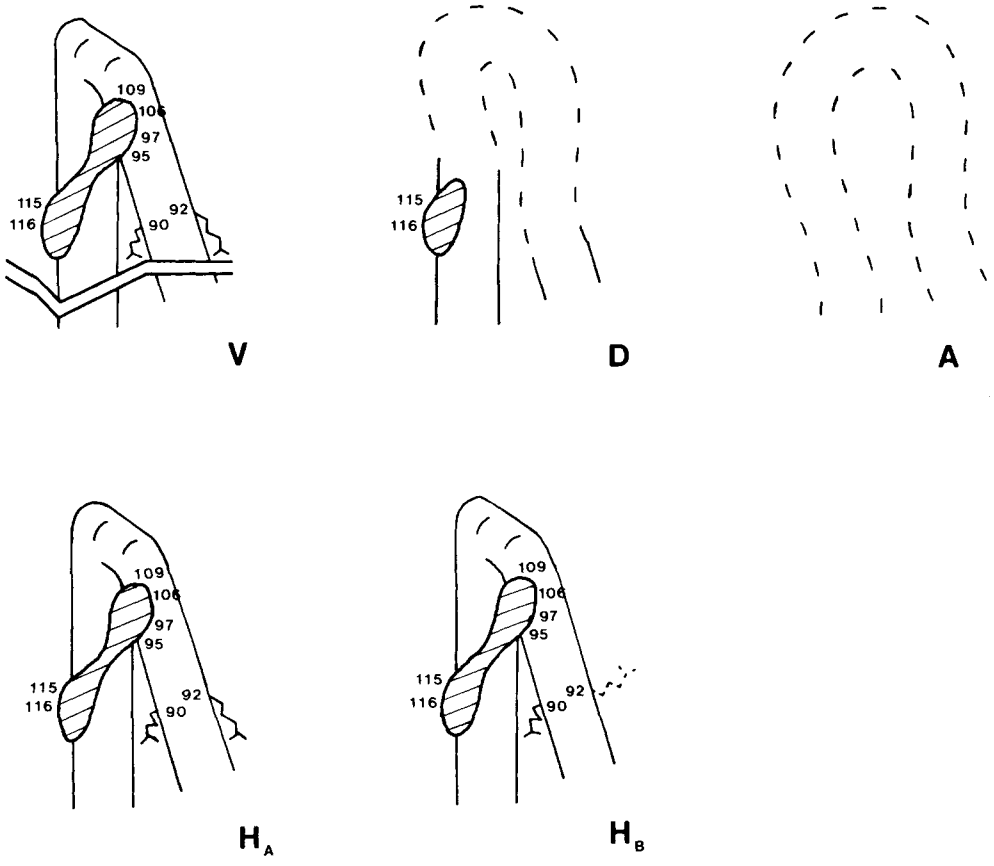

Fig. 2. Schematic representation of the central loop (residues 85–120) in the TMV aggregates. Letters indicate aggregate: Virus, disk, A-protein (speculative), helix A, helix B. Numbers indicate approximate residue locations. Solid lines: Ordered structure. Broken lines: Disordered structure. Shading: "carboxyl cage."

of the radial helices (LR and RR) of the virus extend in beyond the RNA, and are connected by a length of chain which is also partly α-helical (V) [9]. V forms a wall which lines the central hole of the cirus, shielding the RNA from the solvent. It has been suggested [11, 16, 18] that this part of the protein is flexible in the disk in order to allow the RNA to reach its binding site from the central hole.

The packing of the subunits is different in the disk and the virus. The side-to-side interactions are very similar: Two polar regions alternate with two hydrophobic regions [10]. A salt bridge between Arg 122 and Asp 88 has been seen in both structures [9, 10]. However, the interaction in the axial direction (between the rings) of the disk is quite different from the interaction between successive turns of the virus helix [9, 11]. In the virus, the four approximately radial α-helices are closely packed, even between subunits, with the LR helix above and between the RS helix from one subunit and the LS helix from another [11: Fig. 3]. In the disk, one layer is displaced laterally about half a subunit, so that LR is now approximately between LS and RS of the *same* subunit below, although intersubunit packing of helices is now much looser. The molecular details of this

packing have been established [10]: There is a complex network of salt bridges and hydrogen bonds between the subunits. Looking at the 4 Å map of the virus, it appears possible that the inner end of this network (involving part of RS and part of LR) may be maintained by changing the conformation of the side chains, but the rest of the network must be completely different. More details must await a higher resolution map of the virus.

Comparing the protein helices with the virus, we find much more similarity than exists between the disk and the virus. All the significant changes which we can see at this resolution are in or near the RNA binding site. Once again the RNA is absent, but at least one of the three phosphate groups of the RNA is replaced by an anion, probably forming an ion pair with Arg 90. The hydrophobic base binding site on the LR helix is exposed to the solvent, as it is in the disk. The structure of the inner protein, which is disordered in the disk, is intact in both forms of protein helix. This region contains most of the carboxyl cage, and while there are some changes around these carboxyl groups, it must be borne in mind that the virus structure was determined at pH 8, well above the pKs of all the carboxyl groups, whereas the protein helix structures were determined at pH 5.5, where some of the groups would be protonated [1].

Only one significant difference has been found between the A form of the protein helix and the B form. There is a peak of electron density near the inner end of RR which is present in A, but absent in B. This difference was observed regardless of whether this peak was included in the calculation of phases from the virus map. In the virus model [9] the peak has been tentatively attributed to the guanidinium group of Arg 92, which forms an ion pair with an RNA phosphate. We might thus suggest that this arginine retains its conformation in the virus-like A form, but is disordered or substantially moved in the B form. It could bind to an anion, or interact with a nearby carboxyl group – the maps do not allow us to make a definite statement.

DISCUSSION

The transitions between forms of TMV protein all appear to involve conformational changes of the charged groups in the low radius region: The six carboxyls, two of the phosphate-binding arginines (90 and 92), and possibly the other arginines in this region: 41, 112, and 113. The importance of these charged groups is not surprising, since the main parameters affecting aggregation state are pH and ionic strength [2]. Arg 41 is the only arginine in the inner part of the protein for which the main chain is ordered in all forms so far examined. It is possible, however, that the side chain of this residue is more disordered in the protein helical form than it is in the virus, and it may have different conformations in the virus and the disk. The maps are difficult to interpret in this region. Arg 112 and Arg 113 are near the carboxyl cage [9] and probably form ion pairs with two of the carboxylic acids. It is not possible to say whether they change conformation between the virus helix and the protein helix forms. They are part of the flexible loop in the disk, and are therefore disordered in this aggregate, as are Arg 90 and Arg 92. Arg 90 has the same conformation in the virus and both protein forms, binding a phosphate of the RNA in the virus, and an anion in the protein forms. Arg 92 has an ordered conformation in the virus and the A form of the protein helix, binding RNA in the virus and perhaps an anion in the phosphate site in the protein. However, it has a different conformation in the B protein helix. This is the only major difference so far observed between the A and B protein helix forms and may well be the cause of the difference between the forms. Arg 92 is in the side-to-side intersubunit boundary, and a conformational difference

which affected the intersubunit spacing at this point would change the number of subunits per turn in the helix.

The carboxyl residues (the carboxyl cage) are the source of the anomalous pK values near 7 observed by Caspar [1]. Such pKs arise when carboxyl groups are forced into close proximity by the protein structure. The resulting structure can serve as a switch, active under physiological conditions, between different states of the protein [9, 19]. The virus and protein helices have two or three anomalous pKs [1, 12], and in these structures all six carboxyl residues are held in fixed conformations. In the disk, only residues 115 and 116 of the carboxyl cage retain any trace of rigid conformation [10]. This is because, in the absence of the binding energy of the RNA (as in the virus) or the neutralization of some of the carboxyl groups at lower pH (as in the protein helices), electrostatic repulsion within the flexible loop is too great to allow it to fold [9]. This is consistent with the observation [12] that there is only one anomalous pK value in the disk form, presumably arising from these two residues. Since the A-protein (the small oligomer mixture) has no anomalous pK values [12], we might expect the extent of disorder to be even greater in this form. Asp 115 and Asp 116 are just beyond the end of the flexible loop, at the start of the LR helix in the disk, but LR extends further back in the virus. It may unwind even more in the A-protein than it does in the disk, allowing these two residues to lose their fixed conformation.

Comparing all the major forms of TMV protein so far examined, we see that almost all the differences occur in the RNA binding site and the protein region inside this site. This region, containing the carboxyl cage and the V helix, is where control of assembly is believed to reside [9]. The stability of its secondary and tertiary structure appears to depend on the stability of the carboxyl cage: Four of the carboxyls (95, 97, 106, and 109) are actually part of the flexible loop. Since it is ordered in all helical forms, and probably not in any other form, it seems likely that the formation of this central ordered protein region is an essential step in the transition from 17-fold (disk) symmetry to helical symmetry. V helices from neighboring subunits are only 9 Å apart, that is, they are close-packed, so the interactions between neighboring V helices could well determine the aggregation state. Thus we see that if the carboxyl cage acts as a switch [19] between the disk and helical forms, the transition is probably mediated through the formation of the ordered central protein region.

ACKNOWLEDGMENTS

The authors would like to thank Drs. D.L.D. Caspar and K.C. Holmes for valuable discussions during this work, and Drs. L. Makoswski and C.V. Stauffacher for helpful criticisms of the manuscript.

REFERENCES

1. Caspar DLD: Adv Protein Chem 18:37, 1963.
2. Durham ACH, Klug A: J Mol Biol 67:315, 1972.
3. Finch JT, Leberman R, Chang Y-S, Klug A: Nature 212:349, 1966.
4. Finch JT, Klug A: Phil Trans Roy Soc B261:211, 1971.
5. Schramm G: Z Naturforsch 2b:249, 1947.
6. Franklin RE: Nature 177:928, 1956.
7. Mandelkow E, Holmes KC, Gallwitz U: J Mol Biol 102:265, 1976.

8. Gregory J, Holmes KC: J Mol Biol 13:796, 1965.
9. Stubbs GJ, Warren SG, Holmes KC: Nature 267:216, 1977.
10. Bloomer AC, Champness JN, Bricogne G, Staden R, Klug A: Nature 276:362, 1978.
11. Champness JN, Bloomer AC, Bricogne G, Butler PJG, Klug A: Nature 259:20, 1976.
12. Durham ACH, Vogel D, deMarcillac GD: Eur J Biochem 79:151, 1979.
13. Matthews BW, Colman PM, Jansonius JN, Titani K, Walsh KA, Neurath H: Nature New Biol 238:41, 1972.
14. Hardman KD, Ainsworth CF: Biochemistry 11:4910, 1972.
15. Edelman GM, Cunningham BA, Reeke GN, Becker JW, Waxdal MJ, Wang JL: Proc Natl Acad Sci USA 69:2580, 1972.
16. Jardetzky O, Akasaka K, Vogel D, Morris S, Holmes KC: Nature 273:564, 1978.
17. Holmes KC, Stubbs GJ, Mandelkow E, Gallwitz U: Nature 254:192, 1975.
18. Butler PJG, Finch JT, Zimmern D: Nature 265:217, 1977.
19. Caspar DLD: In Markham R, Horne RW (eds): "Proceedings of the Third John Innes Symposium." Amsterdam: North-Holland, 1976, p 85.