Acta Cryst. (1990). A46, 993-997

Fiber Diffraction Analysis of Cucumber Green Mottle Mosaic Virus Using Limited Numbers of Heavy-Atom Derivatives

BY SHARON LOBERT* AND GERALD STUBBS

Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA

(*Received* 28 *March* 1990; *accepted* 30 *July* 1990)

Abstract

The structure of cucumber green mottle mosaic virus has been determined from fiber diffraction data to a resolution of 5 Å, using only two derivatives and without making any specific assumptions about the molecular structure of the virus. Because of the cylindrical averaging of fiber diffraction data, large numbers of heavy-atom derivatives have been required in previous structure determinations, but it is shown that simplifying assumptions about the overlapping intensities in fiber diffraction are sufficient for structure determination at this resolution using only two derivatives.

Introduction

Fiber diffraction methods are used to determine the structures of filamentous macromolecular assemblies that cannot be crystallized. Since the molecular interactions involved in forming these filamentous assemblies are often of primary biological interest, solving the fiber structure is preferable to solving the crystal structure of some fragment of the assembly. Fiber diffraction methods have been successfully applied to nucleic acids and polysaccharides for many years, but structures having larger asymmetric units have been more intractable. Recent developments (Namba & Stubbs, 1985; Stubbs, Namba & Makowski, 1986) have, however, allowed the determination of the structure of tobacco mosaic virus (TMV) at 2·9 Å resolution (Namba, Pattanayek & Stubbs, 1989), and considerable progress has been made with other structures, including the filamentous bacteriophages (Stark, Nambudripad & Makowski, 1990).

The structures of filamentous macromolecular assemblies can be determined from fiber diffraction patterns by a multi-dimensional extension (MDIR) of the protein crystallographic method of isomorphous replacement (Stubbs & Diamond, 1975), but the method requires large numbers of heavy-atom derivatives. This is because fiber diffraction data are cylindrically averaged, so that the diffracted intensity is

$$I(R, l) = \sum_n \mathbf{G}_{n,l}(R)\mathbf{G}^*_{n,1}(R) \qquad (1)$$

(Waser, 1955; Franklin & Klug, 1955), where l is the layer-line number, R is the reciprocal-space radius, and n is the order of the Bessel functions J_n that contribute to the complex Fourier-Bessel structure factor $\mathbf{G}$ (Klug, Crick & Wyckoff, 1958). The number of significant terms contributing to the intensity in (1), denoted here by N, depends on the symmetry and dimensions of the diffracting particle, and on the value of (R, l). For example, for TMV and other members of the tobamovirus group at 2·9 Å resolution, N can be as much as 8. At 5 Å resolution, the maximum value of N for these viruses is 3.

If each $\mathbf{G}$ is known, a Fourier-Bessel transform of the G terms will yield an electron density map, analogous to the electron density map obtained from the F terms for a crystal. The G terms can be obtained by MDIR, sometimes supplemented by data from layer-line splitting (Stubbs & Makowski, 1982; Namba & Stubbs, 1985), but the requirements for numbers of heavy-atom derivatives can be prohibitive.

Namba & Stubbs (1987) showed that the number of derivatives required could be reduced considerably by using assumptions about the relative amplitudes of the Fourier-Bessel structure factors. Overlapping structure factors were assumed to have equal magnitudes, or estimates of their relative magnitudes were taken from a related model structure. These assumptions were applied to data from the native assembly and from one or two heavy-atom derivatives, and conventional single or double isomorphous replacement [see, for example, Blundell & Johnson (1976), p. 364] was used to calculate an initial phase set. This phase set was refined by density modification: the electron density was set to zero in regions considered to represent solvent, and the modified electron density was used to calculate a new set of relative magnitudes for the structure factors. The new estimates for the relative magnitudes of the derivative structure factors took the heavy-atom contributions into consideration. The process can be repeated, but for both single and double isomorphous replacement, the structures converged after a very small number of cycles, usually one or two.

Cucumber green mottle mosaic virus, watermelon strain (CGMMV-W), is a member of the tobamovirus group of plant viruses, of which TMV is the type

* Present address: Department of Biochemistry, University of Mississippi Medical Center, Jackson, MS 39216, USA.

0108-7673/90/120993-05$03.00

member. These are rod-shaped viruses, approximately 3000 Å long and 180 Å in diameter. Identical protein subunits of molecular weight about 17 500 form a helix of pitch about 23 Å, with 49 subunits in three turns. A single strand of RNA follows the basic helix at a radius of 40 Å. The TMV subunit is illustrated in Fig. 1. CGMMV-W has about 36% amino-acid sequence homology with TMV (Meshi, Kiyama, Ohno & Okada, 1983), with most of the nucleic-acid-binding residues and many of the residues making intersubunit charge interactions conserved. Its structure is of considerable interest, however, because two residues that are apparently essential for the disassembly of TMV, Glu 50 and Asp 77 (Namba & Stubbs, 1986), are not conserved. The radial density distribution and the general distribution of intensity in the fiber diffraction pattern (Lobert, Heil, Namba & Stubbs, 1987) suggest that the major structural features of TMV are retained in CGMMV-W.

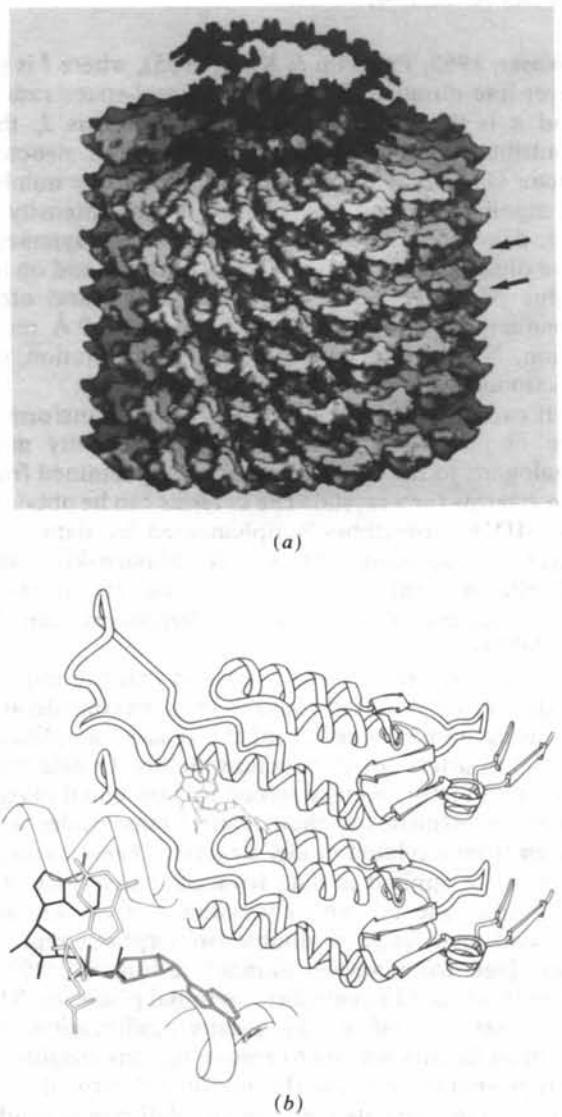


Fig. 1. (*a*) Computer graphics representation of about 1/20th of the TMV particle. Protein subunits are light gray; RNA is dark gray. The RNA is shown extending beyond the end of the protein helix for clarity. Graphics from Namba, Caspar & Stubbs, (1985). (*b*) A ribbon drawing of two subunits of TMV. These two subunits would fit into the right-hand part of (*a*), as indicated by the arrows in (*a*). The core of the monomer is a bundle of four α-helices, designated [in the terminology of Champness, Bloomer, Bricogne, Butler & Klug (1976)] left slewed (LS), right slewed (RS), left radial (LR) and right radial (RR). Three nucleotides of RNA are shown binding between the subunits, and in the enlarged inset.

Table 1. *Heavy-atom parameters*

r, φ, and z are cylindrical coordinates. Units of occupancy (*f*) are on an arbitrary scale.

Ligand	*f*	*r*(Å)	φ(°)	*z*(Å)	*B*(Å²)
MNN	0·3	57·4	7·8	21·6	50
	0·2	42·4	−4·2	32·3	25
Pb^{2+}	0·22	23·4	6·2	22·1	50
	0·45	52·9	3·1	24·7	80

Data collection

The virus was prepared as described by Lobert *et al.* (1987). Oriented gels for diffraction were made by shearing in 0·7 mm quartz capillary tubes (Gregory & Holmes, 1965). X-ray fiber diffraction data were obtained from the native virus, and from two heavy-atom derivatives: methyl mercury nitrate (MMN) in 0·1 *M* sodium phosphate buffer, pH 7·2, MMN to virus coat protein stoichiometry 10 to 1; and lead acetate in 0·075 *M* glycylglycine buffer, pH 8·1, lead to virus coat protein stoichiometry 2·2 to 1. Data to a resolution of 5 Å were recorded, measured and corrected for geometric and other effects as described by Namba & Stubbs (1985). Heavy-atom coordinates and occupancies were determined as described by Lobert *et al.* (1987). The heavy-atom parameters are given in Table 1.

Extraction of Fourier–Bessel structure factors and phase determination

At each point (R, l), the amplitudes of the Fourier–Bessel structure factors (G terms) contributing to $I(R, l)$ in (1) must be determined. Two different sets of initial assumptions about the amplitudes of the G terms were used to compensate for the absence of enough heavy-atom derivatives to determine these amplitudes directly. In the first set ('equal amplitudes'), all amplitudes of significant magnitude were assumed to be equal. [The amplitude of a contributing G term was considered to be significant if the argument of the corresponding Bessel function of order n was greater than $0{\cdot}825(n-4{\cdot}9)$ for an atom at the outer radius of the diffracting particle.] This assumption was made for both the native and derivative data. For the second set ('model amplitudes'), the relative

amplitudes were assumed to be the same as those calculated from the TMV structure. Again, the same relative amplitudes were used for the G terms in the native and the two heavy-atom derivative data sets. It would be possible to include the heavy atoms in the calculation to estimate amplitudes, but this would effectively assume the TMV phases as well as the amplitudes. We preferred to minimize bias toward the TMV structure.

Equal amplitudes have the advantage of completely avoiding bias toward the TMV structure. Model amplitudes probably have the advantage of being more accurate than equal amplitudes.

Having thus decomposed the observed intensities into contributions from the different G terms, double isomorphous replacement was used to calculate phases. Electron density calculated from these phases and amplitudes was weak and often discontinuous (Fig. 2*a*). Solvent flattening outside a hollow cylindrical envelope ['box function' refinement (Makowski, 1981, 1982)] was used to improve the maps. In this procedure, the electron density map [or, in practice, the Fourier–Bessel components $\mathbf{g}_{n,l}(r)$ of the map] is set to zero outside the known inner and outer radii for the virus particle, and back-transformed to give $\{\mathbf{G}\}$. The relative amplitudes and phases of this $\{\mathbf{G}\}$ are applied to the observed intensities, and the resulting $\{\mathbf{G}\}$ is transformed to give a new map. The refinement, monitored by the crystallographic R factor between the observed and back-transformed data, converged by the end of five cycles. The electron density in the map after refinement (Fig. 2*b*) was much more interpretable than in the unrefined map. This refinement did not alter the relative intensities of the contributing G terms. Those relative intensities could be refined by iterations of isomorphous replacement, and that procedure was useful in the higher-resolution work of Namba & Stubbs (1987), but it did not improve the electron density in the present work.

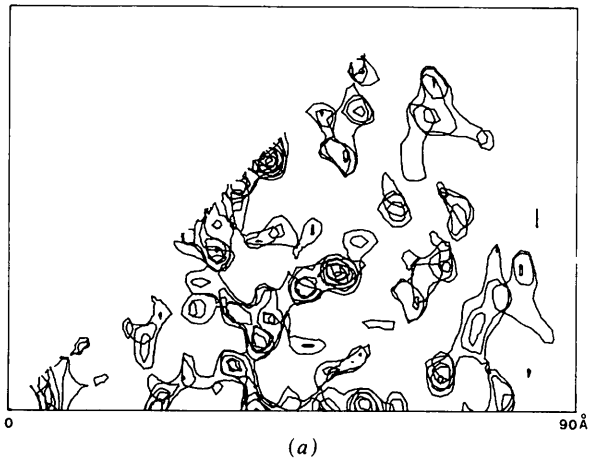


Fig. 2. Three superimposed sections of the CGMMV-W electron density map, orthogonal to the viral axis, at 5 Å resolution. The viral axis is in the bottom left corner; a 45° sector, covering sections through two adjacent subunits, is shown. Part of one of the α-helices can be seen running radially through the center of the maps. (Equivalent helices in the adjacent subunits are partly visible at the edges of the sectors.) (*a*) Before solvent flattening. (*b*) After five cycles of radially constrained solvent flattening refinement. The refinement has clearly improved contrast and continuity.

Model building

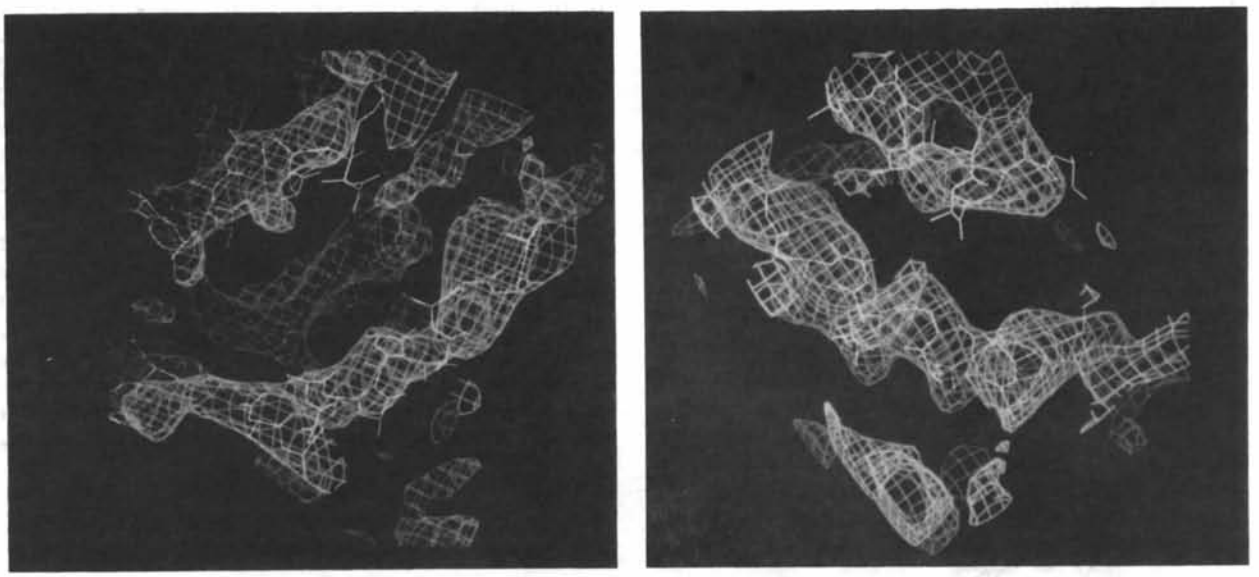
With the TMV structure as a guide, models were built using the graphics program *FRODO* (Jones, 1982) in an Evans and Sutherland PS340 computer graphics system. At 5 Å resolution, only the α-helical regions could be reliably traced, but these regions included about half of the 160 amino acids, and clearly showed the bundle of four α-helices and the carboxy-terminal helix. Polyalanine models fitted to maps calculated using both sets of initial assumptions are shown in Fig. 3.

The dimensions and folding of the protein subunit are very similar in TMV and CGMMV-W (Figs. 1*b* and 3). In both models of CGMMV-W, the top and bottom layers in the four-α-helix bundle are 1–2 Å farther apart than in TMV. This is consistent with the observation that amino-acid volumes (Chothia, 1975) between the layers, and particularly between the left slewed and the left radial helices, are slightly larger in CGMMV-W than in TMV (Lobert, 1989). The helix pitch of CGMMV-W is 23·6(3) Å. This is only slightly greater than that of TMV, so the intersubunit distance between the layers is correspondingly less than in TMV. Apart from these differences, the packing of the α-helices in the two viruses is essentially the same.

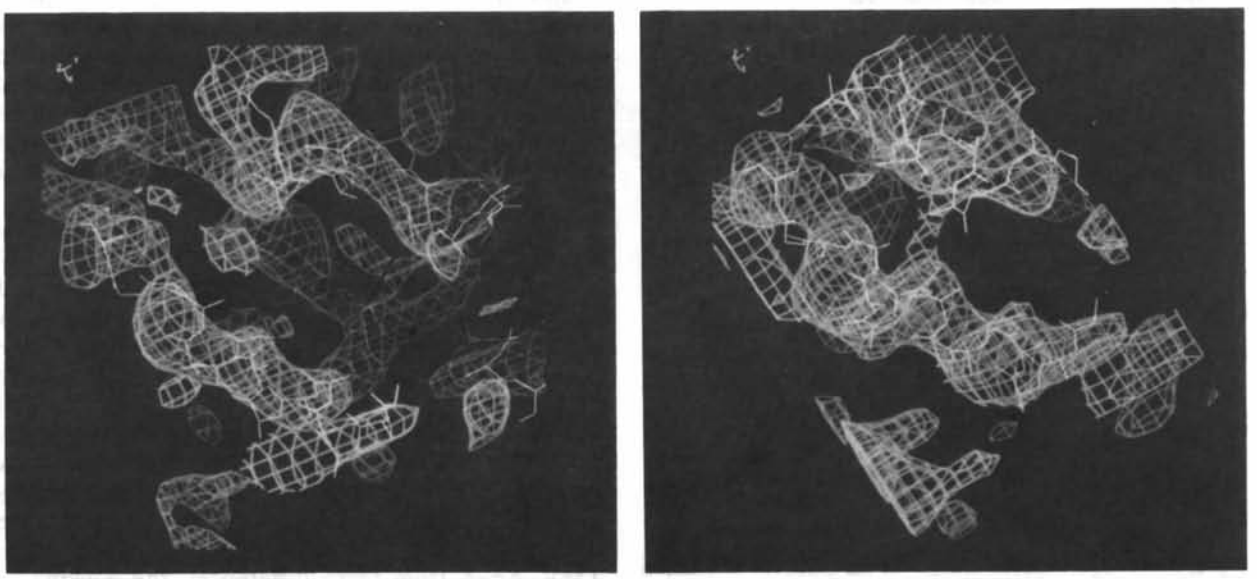
Discussion

The two methods used to estimate structure-factor amplitudes resulted in nearly identical electron density maps (Fig. 3). It is thus evident that, at this

resolution, the assumption of equal amplitudes is sufficient to obtain interpretable maps. This assumption has the advantage of being free of any bias toward a model structure. It is also worth noting that since the observed differences between TMV and CGMMV-W were found using both methods, the use of model amplitude estimates did not noticeably bias the electron density map. Working at higher resolution, Namba & Stubbs (1987) did not obtain maps of quality comparable to those calculated here using equal-amplitude assumptions. It appears, therefore, that, for medium-resolution work (or, more precisely,



(*a*)



(*b*)

Fig. 3. Electron density maps at 5 Å resolution with the CGMMV-W model superimposed, viewed parallel to the viral axis (looking from the top relative to Fig. 1). The axis is in the top left corner. Three α-helices are shown in the maps on the left: from top to bottom, RS, RR and LR (see Fig. 1 for helix nomenclature). Two α-helices are shown in the maps on the right: from top to bottom, RS and LS. (*a*) Calculated from the initial assumption that all significant amplitudes contributing to an intensity are equal. (*b*) Calculated from the initial assumption that amplitudes contributing to an intensity have the same relative magnitudes as those calculated from the TMV model.

in cases where only two or three Fourier–Bessel terms overlap to produce the observed intensity), the equal-amplitude assumption can be of great value in fiber diffraction analysis, both to be sure of obtaining an unbiased map, and to obtain a map when no reasonable preliminary model is available. In view of the difficulties often experienced when attempting to make heavy-atom derivatives of fibrous assemblies, this conclusion has great potential value in studies of viruses, cytoskeletal elements, and other biological filaments.

The structures of TMV and CGMMV-W are clearly very similar, at least in the α-helical core. At 5 Å resolution, it is not possible to say how CGMMV-W compensates, if at all, for the loss of Glu 50 and Asp 77; however, the overall similarity of the two viruses permits us to speculate that CGMMV-W may contain another pair of carboxylates serving the same function. Any of the residues Asp 42, Glu 46, Asp 126 and Glu 130 might be involved in this, but higher-resolution structural analysis will be necessary before any definitive statement can be made.

We thank Rekha Pattanayek for valuable discussions. This work was supported by NIH grant GM33265 and NSF grants BBS-87179149 and BBS-8607624.

References

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography.* New York: Academic Press.
CHAMPNESS, J. N., BLOOMER, A. C., BRICOGNE, G., BUTLER, P. J. G. & KLUG, A. (1976). *Nature (London)*, **259**, 20-24.
CHOTHIA, C. (1975). *Nature (London)*, **254**, 304-308.
FRANKLIN, R. E. & KLUG, A. (1955). *Acta Cryst.* **8**, 777-780.
GREGORY, J. & HOLMES, K. C. (1965). *J. Mol. Biol.* **13**, 796-801.
JONES, T. A. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 303-317. Oxford Univ. Press.
KLUG, A., CRICK, F. H. C. & WYCKOFF, H. W. (1958). *Acta Cryst.* **11**, 199-213.
LOBERT, S. (1989). PhD thesis. Vanderbilt Univ., Nashville, TN, USA.
LOBERT, S., HEIL, P., NAMBA, K. & STUBBS, G. (1987). *J. Mol. Biol.* **196**, 935-938.
MAKOWSKI, L. (1981). *J. Appl. Cryst.* **14**, 160-168.
MAKOWSKI, L. (1982). *J. Appl. Cryst.* **15**, 546-557.
MESHI, T., KIYAMA, R., OHNO, T. & OKADA, Y. (1983). *Virology*, **127**, 54-64.
NAMBA, K., CASPAR, D. L. D. & STUBBS, G. (1985). *Science*, **227**, 773-776.
NAMBA, K., PATTANAYEK, R. & STUBBS, G. (1989). *J. Mol. Biol.* **208**, 307-325.
NAMBA, K. & STUBBS, G. (1985). *Acta Cryst.* A**41**, 252-262.
NAMBA, K. & STUBBS, G. (1986). *Science*, **231**, 1401-1406.
NAMBA, K. & STUBBS, G. (1987). *Acta Cryst.* A**43**, 64-69.
STARK, W., NAMBUDRIPAD, R. & MAKOWSKI, L. (1990). Submitted.
STUBBS, G. & DIAMOND, R. (1975). *Acta Cryst.* A**31**, 709-718.
STUBBS, G. & MAKOWSKI, L. (1982). *Acta Cryst.* A**38**, 417-425.
STUBBS, G., NAMBA, K. & MAKOWSKI, L. (1986). *Biophys. J.* **49**, 58-60.
WASER, J. (1955). *Acta Cryst.* **8**, 142-150.

International Union of Crystallography

Acta Cryst. (1990). A**46**, 997-998

Prices of *Acta Crystallographica* and *Journal of Applied Crystallography*

The Executive Committee of the International Union of Crystallography has determined the following subscription rates and prices of back numbers for *Acta Crystallographica* and *Journal of Applied Crystallography* as from 1 January 1991.

Acta Crystallographica

The following rates will apply for Volumes A47, B47 and C47 (1991). All subscription rates are fixed in Danish kroner. The US dollar equivalents are no longer given because of rapid fluctuations in exchange rates.

Complete volumes, regular price per volume

Sections A, B & C (combined subscription)	Dkr 6700
Section A only	Dkr 1610
Section B only	Dkr 1610
Section C only	Dkr 3800

Complete volumes, reduced price for individuals

Sections A, B & C (combined subscription)	Dkr 1840
Section A only	Dkr 450
Section B only	Dkr 450
Section C only	Dkr 1080

All subscribers in the USA, Canada and Japan should add to the above subscription rates the additional charges for airfreighting as mentioned below.

The reduced-rate subscriptions are ordinarily only available to members of recognized scientific societies, and applications must be accompanied by a written undertaking that the journal is for the personal use of the subscriber and will not be made available to libraries, institutions *etc.* These conditions also apply to persons wishing to order back numbers at the reduced rates.

Single parts

The price of single parts of any Section of Volume 47 (1991) is Dkr 370.

Journal of Applied Crystallography

The following rates will apply for Volume 24 (1991). All subscription rates are fixed in Danish kroner. The US dollar