PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Structure and Assembly of Filamentous Bacterial Viruses

D. A. Marvin and E. J. Wachtel

Phil. Trans. R. Soc. Lond. B 1976 **276**, 81-98
doi: 10.1098/rstb.1976.0099

Phil. Trans. R. Soc. Lond. B. **276**, 81–98 (1976) [81]
Printed in Great Britain

Structure and assembly of filamentous bacterial viruses

By D. A. Marvin† and E. J. Wachtel
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, *U.S.A.*

[Plate 18]

Filamentous bacterial viruses are flexible nucleoprotein rods, about 6 nm in diameter by 1000–2000 nm in length (depending on the virus strain). A protein shell encloses a central core of single-stranded circular DNA. The coat protein subunits forming the shell are largely α-helix, elongated in an axial direction, and also sloping radially, so as to overlap each other and give an arrangement of subunits reminiscent of scales on a fish. This arrangement of α-helices is rather like some models of myosin filaments.

An early step in assembly of the virion is the formation of a complex between the viral DNA and an intracellular packaging protein that is not found in completed virions. Newly synthesized coat protein becomes associated with the plasma membrane of the cell. During the final steps of assembly, the packaging protein is displaced from the DNA and replaced by coat protein as the virion passes out through the plasma membrane of the host cell.

Introduction

Several strains of filamentous bacterial viruses have been identified, all of which infect Gram-negative bacteria. All strains are elongated nucleoproteins measuring about 6 nm in diameter by 1000–2000 nm long (depending on the strain). The protein coat comprises a major protein (about 99 % by mass of the coat) having a molecular mass of about 5000, and a minor protein (about 1 % of the coat) having a molecular mass of about 55 000. The minor protein is required for adsorption of the virion to the host and for penetration of viral nucleic acid into the host. The nucleic acid of the virus is a single-stranded circular DNA molecule, making up less than 12 % of the virion by mass. The DNA is encapsulated at the core of a hollow cylindrical protein shell. The major coat protein molecules can be represented as single rods of α-helix, running roughly parallel to the long axis of the virion. These viruses are therefore interesting models for assemblies of α-helices, such as myosin filaments, or indeed for any assemblies of elongated subunits.

Assembly of the virion takes place in several stages. The DNA is first encapsulated in an elongated complex with the viral gene 5 protein, an intracellular packaging protein distinct from the viral coat protein. As the coat protein is made, it becomes associated with the plasma membrane of the bacterium. Assembly of the completed virion appears to involve displacement of the gene 5 protein from the DNA by the coat protein as the assembling virion is extruded through the plasma membrane. During assembly there is probably a change in DNA conformation, since the DNA-gene 5 protein complex is both longer and fatter than the virion (see A. Kornberg 1974 for review).

Two similar but distinct structural classes of filamentous bacterial virus have been identified from their X-ray diffraction patterns. Class I includes the fd, f1 and M13 strains, the first to be discovered and the best-studied with regard to intracellular assembly processes. Diffraction

† Present address: EMBL, Postfach 10.2209, Heidelberg, Germany.

patterns of the class I type are difficult to analyse because the helical virion is slightly perturbed. Therefore the best-studied structure is the class II structure, exemplified by the Pf1 strain. In the following we shall discuss the current status of the Pf1 structure determination and then use this structural information to interpret experiments on the assembly of class I virions.

Structure of Pf1

General description

A fibre diffraction pattern of a helical assembly of subunits gives two kinds of information. The position of the layer lines, together with the position of intensity on these layer lines, gives information about the arrangement of subunits in the helix, the helix parameters. The overall distribution of diffracted intensity on these layer lines gives information about the shape of the individual subunits. If the major contribution to the intensity on each layer line arises from a single Bessel function, information about three-dimensional structure is available from the two-dimensional fibre pattern.

Analysis of the layer line positions on X-ray fibre patterns of Pf1 indicates that the protein subunits in Pf1 are arranged in a basic helix of pitch *ca.* 1.5 nm with $5s+2$ (s any integer) subunits in 5 turns of the helix. Analysis of the overall distribution of intensity on the diffraction pattern shows that the subunits can be represented as rods of electron density about 7 nm long, overlapping each other, slewing round the helix axis and sloping radially to give an assembly of rods reminiscent of scales on a fish. The sense of the basic helix is the same as the sense of the helix followed by the individual subunits (Marvin, Wiseman & Wachtel 1974*b*).

The X-ray fibre pattern does not directly give the number s in the above equation, since the meridional reflexion corresponding to the axial repeat of the subunit cannot be unequivocally identified. Several indirect arguments from the diffraction data suggest that $s = 4$: That is, there are 22 units in 5 turns of the virus helix (Marvin *et al.* 1974*b*). One can also use density and water content measurements to calculate an axial spacing of 0.33 nm between fd structure units of 6000 molecular mass (Marvin 1966; Marvin & Hohn 1969). Corresponding measurements have not yet been made for Pf1. However we can calculate the relative volumes of the Pf1 and fd coat proteins using the amino acid compositions (Nakashima, Wiseman, Konigsberg & Marvin 1975) and the volumes occupied by the amino acids (Chothia 1975). The relative cross sectional areas of the Pf1 and fd unit cells are known from X-ray measurements (Marvin *et al.* 1974*b*). If the general design of fd is similar to that of Pf1 (Marvin *et al.* 1974*a*), these comparative measurements can be used to calculate that the axial spacing between Pf1 subunits is 0.31 nm. The predicted axial spacing is 0.33 nm for 22 units or 0.27 nm for 27 units in the 7.2 nm axial repeat of a dry fibre. Since density and water content measurements generally overestimate the contents of the unit cell (Langridge *et al.* 1960*b*) as a consequence of the experimental difficulties in eliminating all water from a protein (deBoer 1968; Blaurock 1975), this line of reasoning supports 22 units in 5 turns. Some doubt about the number of units per turn may remain until the completion of the structure analysis, but the general conclusions about the structure will be virtually the same whether s is 4 or 5.

To analyse the structure we apply the classical trial-and-error methods (Lipson & Cochran 1953) that were used to determine the structure of DNA (Langridge *et al.* 1960*a*). On the basis of a general consideration of the diffraction pattern together with chemical data we build

sterically feasible molecular models of the protein shell, calculate their Fourier transforms, and then repeatedly modify the model within the steric constraints so as to improve the fit between the calculated and observed transforms. In our initial studies, we have not treated the DNA in detail.

The Pf1 protein subunit has concentrations of acidic residues at the *N*-terminus, hydrophobic residues in the middle, and basic residues at the *C*-terminus (Nakashima *et al.* 1975). The coat protein is largely α-helix (Day 1966, 1969; Thomas & Murphy 1975; Green & Flanagan 1976), so for convenience we represent each coat protein molecule as a single gently curved length of α-helix. In fact there may be some non-helical regions in the protein. We orient this α-helix length so that the acidic residues are directed outwards, where they would explain the low isoelectric point of the virion, and the basic residues are directed inwards, where they would neutralize the DNA phosphates in the virus core. The hydrophobic residues in the centre of the sequence would then interact with hydrophobic residues on neighbouring proteins. The importance of hydrophobic bonds between the subunits is indicated by sensitivity of the virus structure to disruption by organic solvents, and resistance of the structure to extremes of ionic strength and pH (Marvin & Hohn 1969). The completed protein shell is roughly cylindrical, with about 1 nm inner radius and 2.8 nm outer radius, with a basic inner surface and an acidic outer surface sandwiching a hydrophobic cylindrical shell centred at about 1.7–2 nm radius. This hydrophobic shell is seen on the radial electron density distribution as a trough having an average electron density about equal to that of water (Wachtel *et al.* 1974). The low average density is caused by the fact that the density of the largely hydrocarbon side chains in this central shell is substantially below that of water, so that the average electron density, including that of the α-helix backbone, is about that of water (figure 4*b* of Marvin & Wachtel 1975).

In our first calculations of the Fourier transform of the virus model, we used a 46-residue length of polyalanine α-helix to approximate the 46-residue protein. The coordinates of this 46-residue α-helix were generated using the equation of Crick (1953*a*) for the coordinates of atoms in an α-helix that has been curved along a large pitch major helix; but with two modifications. First, to handle the radial overlap of the protein molecules, the radius r of each atom in the major helix frame was increased by a value proportional to its height z above the origin of the α-helix. Secondly, the α-helix was divided into segments (for most calculations we used four segments, between 7 and 15 residues long), each of which had a different major helix pitch and a different $\Delta r/\Delta z$. Therefore, it was possible to build models not only with various general orientations of the α-helix, but also with various degrees of bending within the α-helix.

Stereochemical constraints between neighbouring protein molecules were included by calculating the contact distance and crossing angle between the axes of neighbours. The important contacts within the virus helix are of two types. If we call an arbitrary molecule the '0' molecule, and count from this molecule along the basic helix, we find that the two most significant contacts are those between the 0 and 5 molecules and those between the 0 and 9 molecules. Distances between the axes of the 0 and 5 and the 0 and 9 α-helices were considered optimal at 1 nm, and were permitted as low as 0.8 nm. Values above 1 nm are sterically possible but energetically less favoured since van der Waals attraction between neighbours is expected; high values were therefore considered undesirable. The fit of side chains of one α-helix into the space between side chains on a neighbour was analysed as discussed by Crick (1953*b*). Calculations were made to ensure that the β-carbons on one α-helix were roughly

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

centred in the space between β-carbons on the neighbouring helix over the length of the contact region in our models.

Internal stereochemistry of the α-helix was refined by using the computer program of Diamond (1966) to thread a stereochemically optimal polypeptide chain through a set of guide coordinates. All α-helix backbone atoms (i.e. the atoms of polyglycine) were used as guide coordinates. Refinement was only possible if the internal tetrahedral angle at the α-carbon was permitted to vary. In our refined models this angle was within 5–10° of the optimal value 110°. The Ramachandran torsion angles (ϕ, ψ) in the refined models were within the range found for α-helical residues in globular proteins (Ramachandran & Sasisekharan 1968).

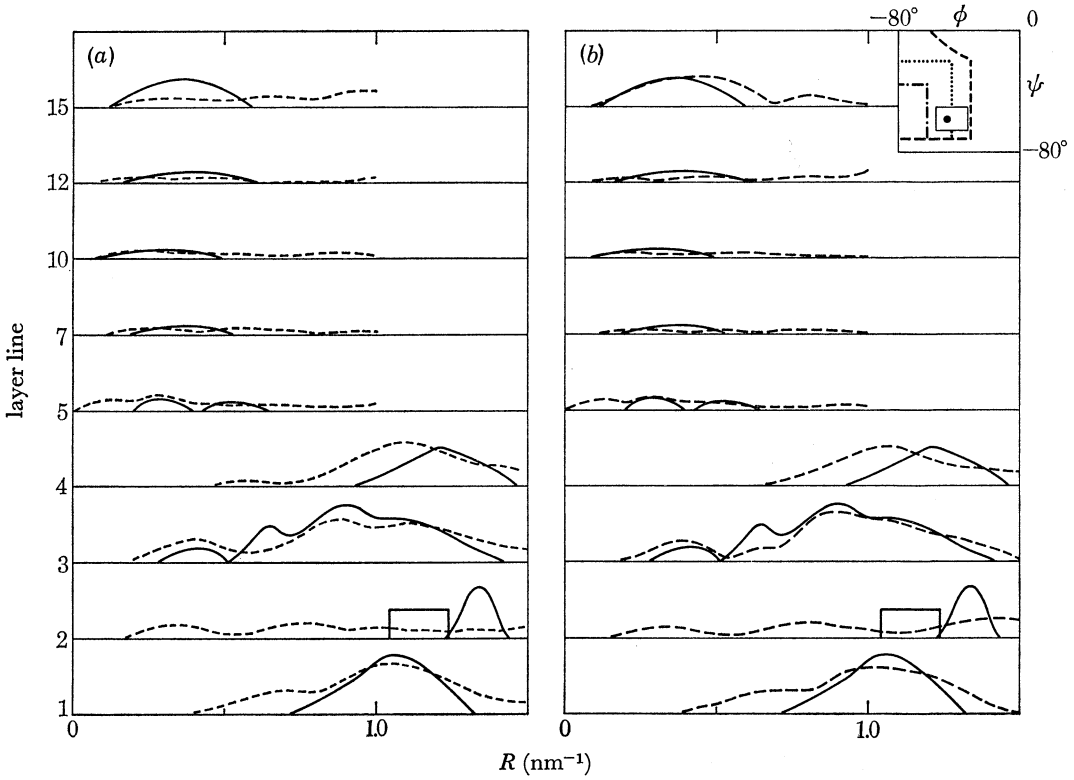


FIGURE 1. Comparison of calculated off-equatorial transform of polyalanine models with observed diffraction amplitudes. The model is a left-handed helix having 22 units in 5 turns, with an axial repeat of 7.7 nm. The helix repeat unit is a 46-residue length of polyalanine α-helix. The transforms were calculated using water-weighted atomic scattering factors. Observed amplitudes, from Marvin & Wachtel (1975), are solid curves; calculated amplitudes are dashed curves. The rectangular box on $l = 2$ represents integrated intensity for which the shape was not determined. (*a*) Transform of the model described by Marvin & Wachtel (1975). (*b*) Transform of a model differing from (*a*) by an axial expansion of the α-helix by 10%, and a rotation of the α-helix by −50° around its own axis. Inset: Distribution of Ramachandran angles for this model. All angles in the model fall within the square box. The dot represents the ideal α-helix torsion angles. The permitted region for torsion angles with values of the internal α-carbon angle of 105°, 110°, 115° are to the left of the dashed lines (reading left to right).

Calculation of diffracted intensity at low resolution (less than about 2 nm^{-1} in reciprocal space) must take into account the fact that in this region the molecules diffract relative to the electron density of the surrounding medium (essentially water) rather than relative to vacuum. A semi-empirical solution to this problem was found by Langridge *et al.* (1960*a*) for

the calculation of DNA transforms. The transform of the molecule imbedded in a sea of water is, by Babinet's principle, identical to the transform of the molecule in vacuum minus the transform of a volume of water having the shape of the volume excluded by the molecule. Langridge *et al.* (1960*a*) approximated this excluded volume by summing a set of overlapping spheres of water, each centred at one of the atoms of the molecule. Thus the transform calculation required only that the normal scattering factors associated with each atom be replaced by a set of water-weighted scattering factors: the normal scattering factors decreased by the transform of an appropriate sphere of water. Since this approximation is quite good for regions of reciprocal space less than 2 nm^{-1} (Ninio, Luzzati & Yaniv 1972; Fedorov, Ptitsyn & Voronin 1974) we used it for most of our trial transform calculations. At various stages during the refinement, we calculated more accurate transforms of models using the method of Fedorov *et al.* (1974). The excluded volume was defined by drawing the van der Waals outline of the model (Lee & Richards 1971). This volume was then approximated by a set of cubes placed on a cubic grid at 0.175 nm intervals. The Fourier transform of the grid of cubes, weighted by the electron density of water, was calculated and subtracted from the transform calculated for the model using unmodified scattering factors.

The use of water-weighted scattering factors with a polyalanine α-helix to approximate the diffraction pattern of the virus implicitly uses the observation of Cohen & Holmes (1963) that the transform of polyalanine is a good approximation to the transform of paramyosin at low resolution. The average electron density of the disordered side chains is not much higher than that of water (Fraser *et al.* 1965), so we can consider the protein to be backbone immersed in a sea of side chains plus water, rather than backbone plus side chains, immersed in a sea of water. Over a hundred models, including a wide range of parameters, were examined during this stage of the structure analysis. The transform of our best model derived by this method (Marvin & Wachtel 1975) is shown in figure 1*a*.

The intensity calculated for this model is generally in agreement with the observed diffraction, but there are two major exceptions. (1) The observed intensity on the fifteenth layer line centred at 0.35 nm^{-1} is substantially larger than that on the fifth through twelfth layer lines, but the calculated is not. (2) The fit of calculated to observed on the equator is poor, and not just because the DNA is not included in the calculation. This can best be seen by comparing the cylindrically averaged radial electron density distribution calculated from observed amplitudes and phases on the equator (Wachtel *et al.* 1974) with the corresponding radial density calculated from the model transform (figure 2).

Fifteenth layer line

The low calculated intensity on the fifteenth layer line for our initial models was unexpected, since the fifteenth layer line corresponds to a spacing of about 0.5 nm, and the α-helix pitch repeat gives strong axial intensity on the α-helix transform at about 0.5 nm. If the repeat of the virus helix and the repeat of the α-helix were not exactly commensurate, the fifteenth layer line might not sample the J_1 maximum of the α-helix transform. We therefore calculated transforms with slight variations of the α-helix parameters. As the α-helix pitch is expanded by a few percent, the calculated intensity on the fifteenth layer line increases substantially (figure 1*b*). The Ramachandran angles of the expanded α-helix are within the permitted limits (figure 1*b* inset), as are the internal tetrahedral angles at the α-carbon.

An elongation of the α-helix by a small percentage is indicated not only by the fifteenth

layer line, but also by the electron density map. Calculation of electron density using the observed amplitudes with the phases of the model (Marvin & Wachtel 1975) shows a rod of electron density that generally follows the model α-helix, but extends a tenth of a nanometer beyond each end of the model.

Equator

In the general description of the Pf1 structure, we explained the minimum in the radial density at about 1.8–2 nm in a qualitative manner as being due to a concentration of low density hydrocarbon side chains in the central portion of the amino acid sequence. It is therefore not surprising that our first approximation to the subunit structure, a uniform polyalanine rod without variation in side chains along the sequence, does not satisfactorily reproduce the equator (figure 2). For the next level of approximation, we included the side chains explicitly in the model. This can conveniently be done using the Diamond (1966) program. Side chains in an energetically favourable conformation were added in the appropriate sequence during refinement of the backbone coordinates, and radial density was calculated. To improve radial density, it was necessary to rotate the subunit around the α-helix axis by $-50°$ with respect to the model of Marvin & Wachtel (1975). This radial density was substantially closer to the observed than the radial density of the model without side chains. We then refined the positions of the side chains to further improve the radial density distribution and the stereochemistry, taking into account the constraints on internal torsion angles along the side chains (Ramachandran & Sasisekharan 1968), and the hard sphere repulsions both within and between the α-helical subunits. The radial density distribution of the current best model of this type is shown in figure 2. The van der Waals outline of a related model is shown in figure 3. The few overlapping atoms are not serious, and may be corrected by further refinement.

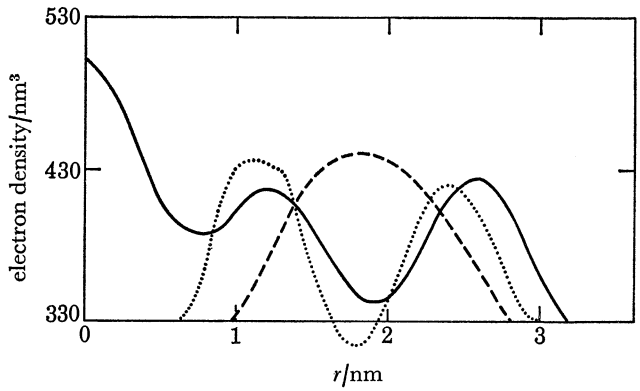


FIGURE 2. Radial electron density distributions. The observed distribution (solid curve) was determined from the amplitudes and signs of the equatorial continuous diffraction data of Pf1 as described by Wachtel *et al.* (1974), except that we used data extending to 0.9 nm^{-1}. A similar determination using Pf1 data from crystalline reflexions gives a similar observed radial density distribution (D. L. D. Caspar, personal communication). Radial densities of the models were calculated from the equatorial transforms of the models to a resolution of 1.0 nm^{-1}. Radial density of the polyalanine model of figure 1*a* is shown dashed. Radial density of the model of figure 1*b* with Pf1 side chains added in a conformation to optimize radial density distribution and van der Waals contacts is shown dotted. For this calculation, water was treated as in figure 4. Axial sections of this model are shown in figure 3.

The Fourier transform of this model (figure 4*a*) has a flaw not present in the transform of the polyalanine model: The calculated intensity on the fifth to twelfth layer lines is too high, when compared with the intensity on the first and third layer lines. Treating the side chains

as a part of the uniform background gave a good off-equatorial transform but a poor equator; including the side chains explicitly gave a poor off-equatorial transform but a good equator. This dilemma could be resolved if there were indeed considerable local disorder in the position of the sidechains, but denser sidechains were nevertheless near the inner and outer surface of the protein shell, with less dense sidechains between. Feasible types of disorder are a time

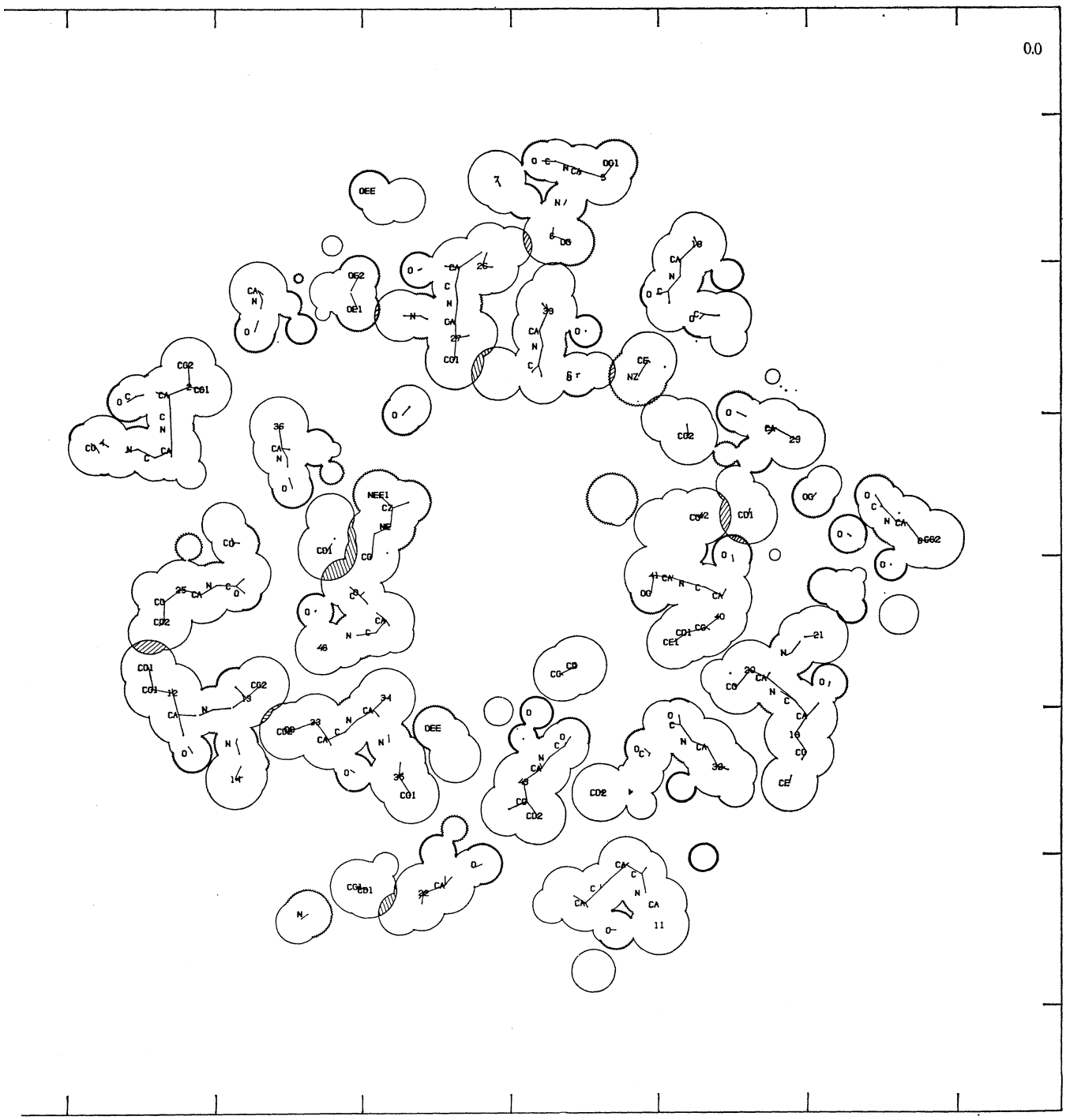


FIGURE 3(*a*). For description see following page.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

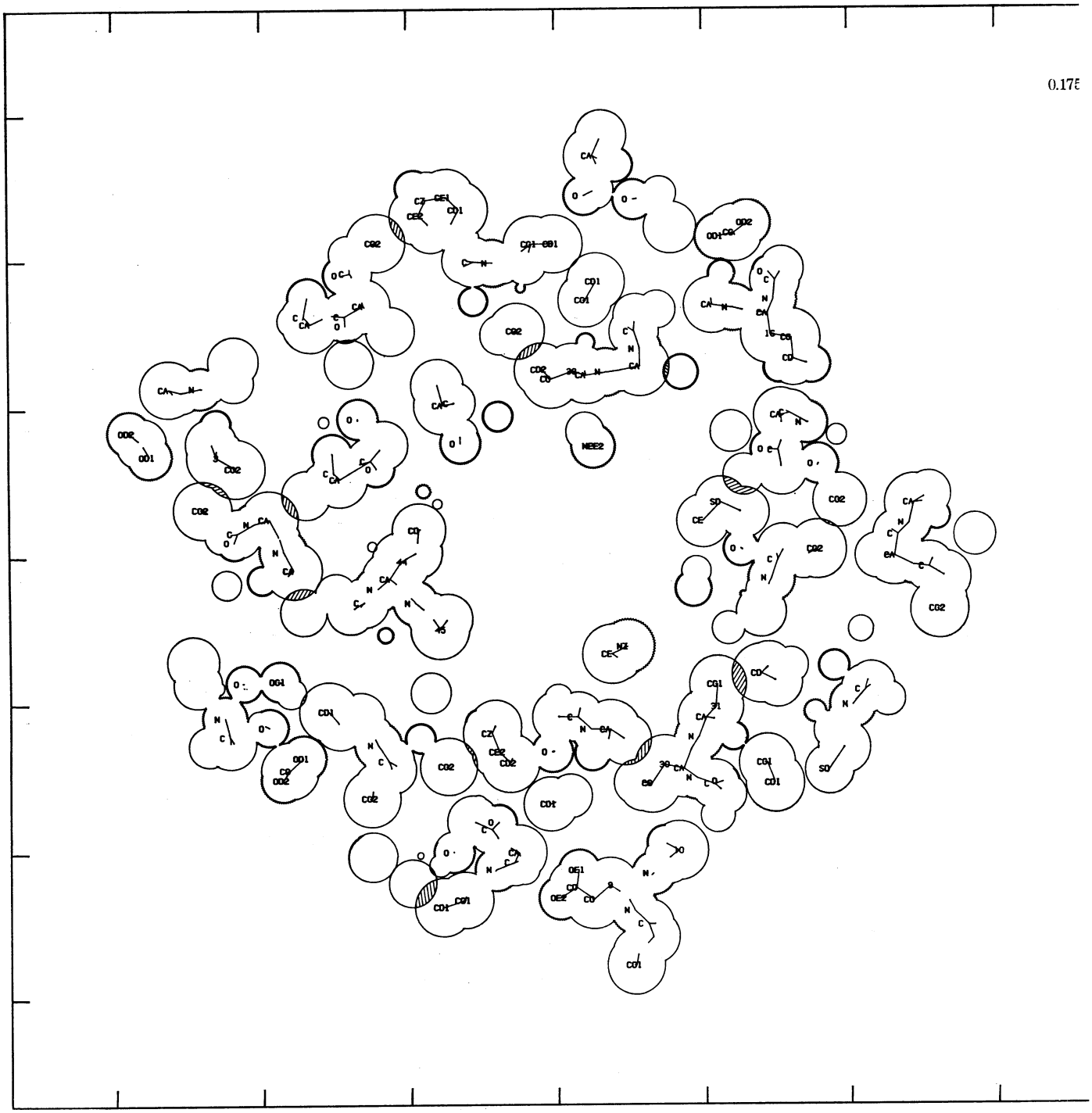


FIGURE 3. Sections through the van der Waals outline of the final model of figure 2, refined by 12 iterations of energy minimization (Levitt & Lifson 1969). The two axial sections are at heights $z = 0$ nm and $z = 0.175$ nm above an arbitrary origin. The section at $z = 0.35$ nm would be identical to the section at $z = 0$ nm, but rotated by $-81.81°$. Sections through the entire virion can be generated by repeated use of this rotation and translation. The outline was drawn using the method of Lee & Richards (1971). Jagged lines outline the hydrophilic N and O atoms. Numerals placed at the β-carbon positions indicate residue number in the amino acid sequence. Letters indicate positions of other atoms. Overlapping van der Waals radii between neighbouring subunits are shaded. Lines connect the centres of all atoms within 0.0875 nm of the plane of the section. Scale divisions are 1 nm.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

average disorder, similar to a temperature factor, or a space average disorder such as screw disorder (Klug, Crick & Wyckoff 1958) of each α-helix along its own axis.

In our first attempt to calculate the transform of such a disordered structure, we replaced all atoms of each sidechain beyond the β-carbon by a sphere at the centroid of the side chain (Levitt & Warshel 1975) having the volume (Chothia 1975) and number of electrons appropriate for the sidechain. This model has a satisfactory equator but unsatisfactory higher layer lines. We then reasoned that since the equatorial transform depends only on radial periodicities, we could cylindrically average the electron density of the sidechains throughout the volume of the model that is not occupied by the polyalanine backbone. This procedure leaves the equatorial distribution the same as that of figure 2 but at the same time returns the off-equatorial layer lines to the desirable state of the uniform background transform (figure 4*b*). The current best model is illustrated in figure 5, plate 18.

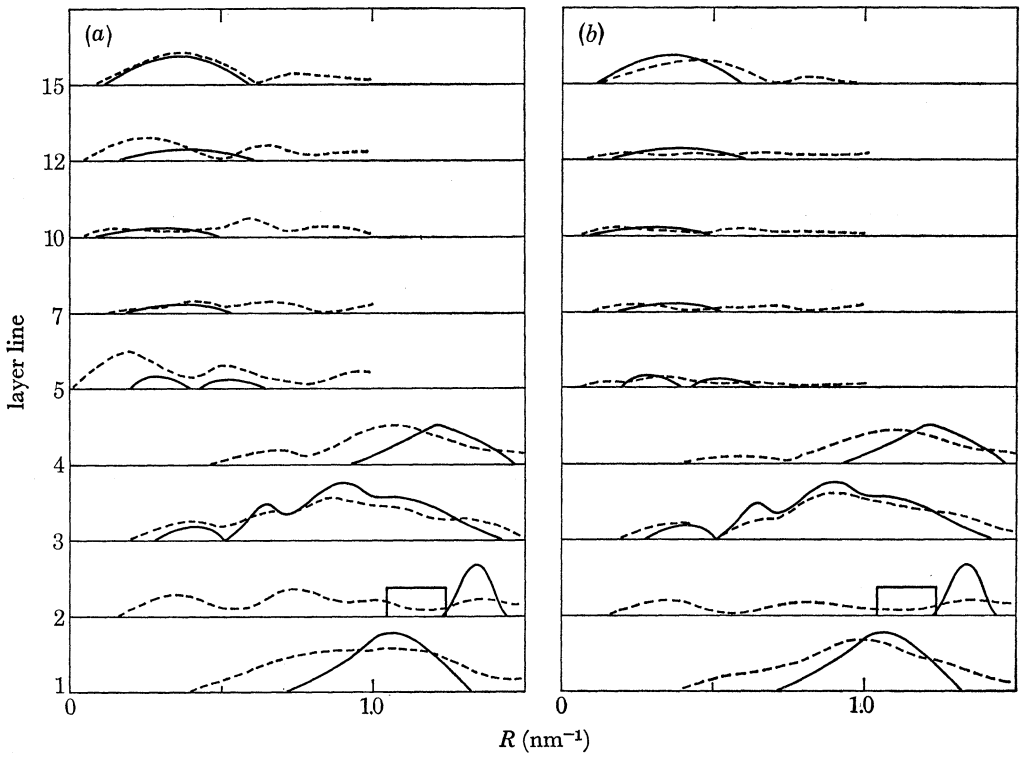


FIGURE 4. Comparison of calculated off-equatorial transforms of Pf1 models with observed diffraction amplitudes. Diffraction from water was treated by the cubic grid method, where the grid was defined by the van der Waals outline of figure 3. No water molecules were included in the interior of the protein shell, even in the six spaces having a volume sufficient to hold a water molecule. The excluded volume of the subunit in this model is about 8% smaller than the 6 nm³ calculated from the amino acid volumes (Chothia 1975). Observed amplitudes, as in figure 1, are solid curves; calculated amplitudes are dashed curves. (*a*) Transform of the model shown in figure 3, with side chains in fixed positions. (*b*) Transform of the same model with cylindrically averaged side chain density.

DNA

The DNA of Pf1 is a single-stranded circular molecule making up about 6% by weight of the virion, or about one nucleotide per coat protein molecule (Wiseman, Berkowitz & Day 1975). Therefore the axial spacing between nucleotides along each strand must be about twice

that between coat protein molecules. Various models for the DNA can be devised, but the small contribution by DNA to the total diffraction pattern of the virion makes it difficult to test DNA models. One useful calculation that is independent of models is the radial density distribution. The difference between the radial density distribution calculated for the final protein model in figure 2 and the observed radial density distribution gives the radial density distribution of the DNA. The DNA is concentrated in a central core.

Symmetrical assemblies of elongated subunits

The model of Pf1 structure shows several symmetry features in addition to $22/\bar{5}$ helical symmetry. These symmetry features can be divided into two types, those concerned with the α-helical subunits as a whole, and those concerned with the helical arrangement of sidechains within the α-helix.

To treat the first type of symmetry we represent each α-helical subunit by a segment of a space curve following the axis of the α-helix. These curves overlap each other so that nearest-neighbours are not adjacent curves in the $22/\bar{5}$ basic helix, but curves in higher turns of this helix. The regions of closest contact between two curves are also at different radii on the two curves. We assume that optimum packing of these curves will involve symmetrical centering of each curve between its nearest neighbours, along the length of the curve, and we look for symmetry principles to describe this packing. Similar questions about packing arise in the study of phyllotaxis, the ordered arrangement of leaves around the stem of a plant (Coxeter 1969; Adler 1974). Phyllotaxis theory predicts that a helix with between 4 and 5 elongated units per turn in the basic helix will have close contacts between the 0 unit and the 4, 5, 9, 14... units, as observed for Pf1.

A convenient way to study the detailed geometrical relationships between the curves is to introduce the concept of the helicoid (Struik 1961). A helicoid is a twisted surface that is generated by the helical motion of a space curve about the helix axis. For each helix through a set of points on the helix net of Pf1 structure, there is a helicoid through the set of curves representing the corresponding subunits. Consider the contacts between the 0 subunit and the 5 and 9 subunits in Pf1. A helicoid of pitch −13 nm (negative sign means left-handed) passes through the curves representing the 0, 5, 10, etc., subunits; and another helicoid of pitch −70 nm passes through the curves representing the 0, 9, 18, etc., subunits. The identical curves on the helicoidal surface are called generators of the helicoid. Curves on the helicoidal surface running orthogonal to the generators are called orthogonal trajectories of the generators. One measure of the distance between closest points on the 0 and 5 curves is the length of the segment of orthogonal trajectory that is cut off by the 0 and 5 generators. One measure of the angle which the 0 and 5 curves make with each other in space is the torsion along the orthogonal trajectory between the two. Neither the distance between the two generators nor

Description of plate 18

Figure 5. The polyalanine model of figure 1*b*. The model consists of a series of 0.35 nm axial sections having the van der Waals outline of the polyalanine backbone. (*a*) Exterior view, showing the overlap of the 0, −5, and −9 subunits (in black). (*b*) Cut-away view, showing the gentle curve of a subunit (black) from large to small radius through the wall of the virion. The central cavity in the model is filled with DNA in the virion.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

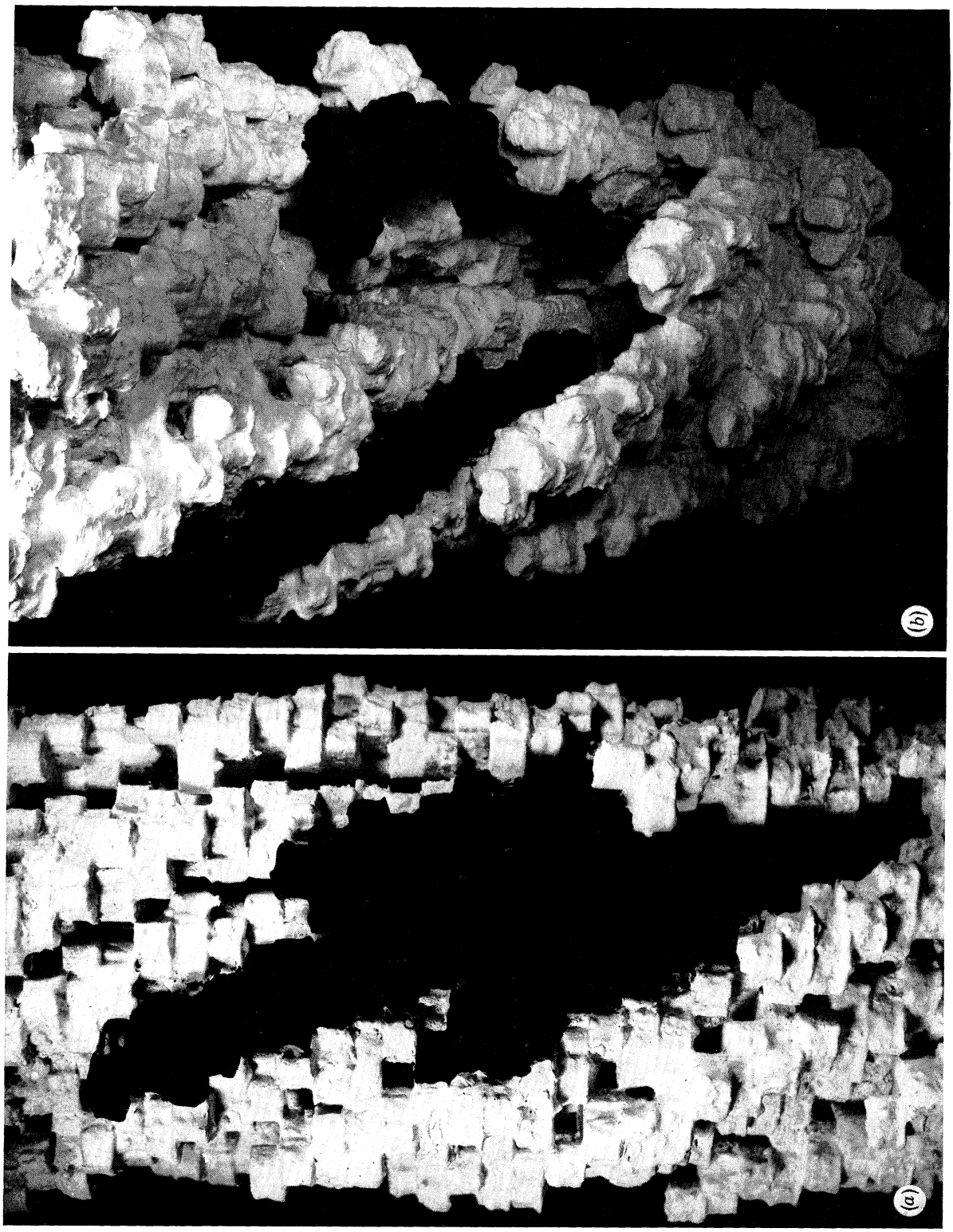


FIGURE 5. For description see opposite.

(*Facing p.* 90)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the torsion along the orthogonal trajectory need be the same at every point along the generator. The way in which they vary along the generator depends both on the equation of the generator and on the pitch of the helicoid. Contacts between curves on the 0–5 helicoid cannot be identical to contacts on the 0–9 helicoid. That is, a simple helical relationship between subunits cannot place the 0 subunit symmetrically between the 5 and 9 subunits. The curve may have an equation that equalizes the distance from 0–5 and 0–9 along the contact length. Even in this case, however, the 0 curve will not be symmetrically placed between the 5 and 9 curves, because the torsion within the 0–5 helicoid is not the same as that within the 0–9 helicoid.

Since a regular helical array of subunits cannot give a symmetrical (minimum energy) conformation, perturbation of the helical array in order to lower the free energy of the assembly is to be expected. Five interleaved helicoids in the 0–5 directions pass through all subunits in the helix. A slight perturbation of the structure by sliding the 0–5 helicoids relative to one another, while retaining regular contacts within these helicoids, could improve the 0–9 contacts. A perturbation of this sort would explain the meridional reflexion on the fifth layer line of Pf1, and a more pronounced perturbation of the same sort would explain the strong 1.6 nm meridional reflexion on the class I diffraction pattern (Marvin *et al.* 1974*a*).

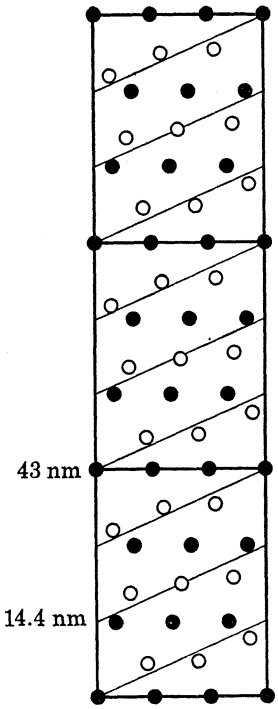


FIGURE 6. Surface lattice of a possible perturbed arrangement of myosin molecules in vertebrate skeletal muscle (open circles). The filled circles show the corresponding lattice proposed for myosin heads by Squire (1973).

The fact that the elongated subunits are themselves helices, and not just smooth rods, introduces a second type of symmetry consideration. The rules governing interaction between neighbouring α-helices have been discussed by Crick (1953*b*). The sidechains on an α-helix form a set of 'knobs', and each space defined by the 0, 3, 4, and 7 sidechains along the basic α-helix form a set of 'holes'. The 0, 7, 14... knobs on one α-helix will fit into corresponding

holes on a roughly parallel neighbour if the helices are twisted slightly around their own axes, from 7.2 residues in two turns for the undistorted α-helix, to 7 residues in two turns. Such a twist becomes sterically feasible if the two helices wind around a common axis, with the axes of the two α-helices making an angle of 20° with one another. The β-carbons on the two neighbours will then be related by a twofold screw down the axis between them, having a pitch of 1.05 nm, the axial spacing for 7 residues. In our case, similar symmetry relationships exist between neighbouring α-helices, but the axis of the twofold screw is a generator on the same helicoid midway between the generators defining the axes of the two α-helices. Therefore the twofold screw axis is both curved in space and skew to the virus axis. This twofold axis does not relate each subunit to its neighbour as a whole, but only refers to local knobs-into-holes packing of β-carbons, irrespective of the detailed nature of the sidechains.

It is interesting to reconsider some models for other assemblies of elongated subunits in the light of these principles. Myosin molecules in striated muscle are thought to be arranged in a three-start helix of pitch 43 nm, with an axial spacing of 14.4 nm between myosin molecules, giving rise to the strong 14.4 nm meridional reflexion (Squire 1973). However the considerations discussed here suggest that the 14.4 nm meridional reflexion could instead indicate a perturbation repeating along a 14.4 nm pitch (figure 6). The true axial repeat would then be about 5 nm if the repeat unit is a dimer of α-helical polypeptide chains, as is commonly believed. It is further possible that the structural repeat unit in the myosin filament is actually a single α-helical polypeptide chain, and dimers are observed in solution because there are strong bonds between α-helices in one direction within the myosin filament that lead to disassembly into pairs of α-helices. In this case there would be a meridional reflexion on diffraction patterns of muscle at about 2.5 nm that might be buried under the 2.6 nm actin reflexion.

Assembly of fd

Membrane proteins

Several experiments indicate that assembly of fd takes place at the bacterial membrane. Progeny virions are released from bacteria without lysing or killing the host (Hoffmann-Berling & Mazé 1964). Pools of viral DNA and viral coat protein but no pool of infectious virions can be detected in the host (Trenkner 1970). One explanation of these experiments is that the coat protein lies across the lipid bilayer of the plasma membrane, with the hydrophobic central region and charged ends of the α-helical rod corresponding to the hydrophobic middle and charged surfaces of the bilayer (Marvin & Hohn 1969); the virion would then be assembled as it is extruded out of the bacterium perpendicular to the plasma membrane. Study of the intracellular location of viral gene products confirms that both the major and the minor coat proteins are associated with the plasma membrane (Smilowitz, Carson & Robbins 1972; Webster & Cashman 1973; Lin & Pratt 1974; A. Kornberg 1974). The viral gene 5 protein (5 protein or 5p) is found largely in intracellular DNA-5p complexes (Pratt, Laws & Griffith 1974), but some 5p is also associated with the membrane (Lin & Pratt 1974). Yet another protein, 4p, is also associated with the membrane, and may be indirectly involved in assembly (Lin & Pratt 1974).

The major and minor coat proteins that are made in an *in vitro* protein synthesizing system are respectively about 6 and 30 amino acids longer than the corresponding proteins isolated from the virion (Konings, Hulsebos & van den Hondel 1975), suggesting that the proteins are

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

cleaved during assembly of the virion. Processing of viral proteins by proteolytic cleavage is often found in virus assembly (Eiserling & Dickson 1972). An accepted explanation for cleavage is that the structure of the uncleaved protein is optimal for the process of assembly, and the structure of the cleaved protein is optimal for stability of the completed assembly. The viral proteins with unidentified functions (1p, 4p, 6p, 7p) are candidates for the cleaving enzymes.

Packaging of the DNA

The initial step in the packaging of fd viral DNA is the formation of the linear intracellular DNA-5p complex. The gene 5 polypeptide chain has a molecular mass of 9689 (Nakashima, Dunker, Marvin & Konigsberg 1974) and appears to exist as dimers in solution (Oey & Knippers 1972; Rasched & Pohl 1974). It binds to single-stranded DNA in the ratio of one 5p molecule for every four nucleotides, perhaps by intercalation of tyrosine side chains between nucleotide bases (Day 1973; Anderson, Nakashima & Coleman 1975). The 5p has both negative and positive roles in DNA replication. On the one hand, 5p appears to inhibit synthesis of a complementary DNA strand on the growing viral single-stranded DNA (Salstrom & Pratt 1971; Mazur & Model 1973; Geider & Kornberg 1974); on the other hand 5p appears to have a positive role in promoting viral DNA synthesis (Staudenbauer & Hofschneider 1973). These apparently contradictory roles can be understood if the growing tail of single-stranded DNA is folded back on itself by interaction with a 5p dimer, and this DNA-5p hairpin grows by sliding along itself as shown in figure 7. The negative function would then be simply steric hindrance of complementary strand synthesis; the positive function might involve active pulling of the growing DNA strand off the rolling circle template by the advancing 5p hairpin. The linear shape of the completed DNA-5p complex suggests that 5p has an important role in packaging the circular DNA into the general elongated form needed for completed virions. The DNA-5p complex measures 1100 nm by 16 nm (Pratt *et al.* 1974) compared with 900 nm by 6 nm for the virion, suggesting that there is a configurational change in the DNA when 5p is replaced by the coat protein. The 5p is not used up during assembly but is repeatedly recycled into subsequent assembly events (Pratt *et al.* 1974).

Molecules of 5p that are complexed with oligonucleotides can be crosslinked into units as large as octamers (Rasched & Pohl 1974). Some electron micrographs of DNA-5p complexes show a beaded appearance, with a bead diameter of about 6 nm (Alberts, Frey & Delius 1972) the diameter expected for an octamer of 5p. There may be parallels between the structure of the DNA-5p complex and the beaded octamer structure of histones in chromatin (R. D. Kornberg 1974). It would be interesting to see if the sequence of 5p, like that of histones, is conserved across strains with substantially different coat protein sequences.

The linear density of DNA in the class II virion Xf is about the same as that of the class I virions such as fd or If1, about 0.3–0.4 nm between nucleotides along each strand of the circular DNA. The linear density of DNA in the class II virion Pf1, on the other hand, is about half that of DNA in Xf and the class I virions (Wiseman *et al.* 1975). The lower linear density of Pf1 DNA is also shown by the axial projection of the central core mass (figure 2), which is lower than that of the If1 core (Wachtel *et al.* 1974). Since the diffraction patterns of Pf1 and Xf are quite similar (Marvin *et al.* 1974*b*) the arrangement of protein subunits must be quite similar. It is therefore unlikely that a detailed relationship exists between the arrangement of protein and the arrangement of DNA in the virion. Instead, DNA and protein probably interact as a generally negatively charged rod and a generally positively charged

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

sheath, respectively. All detailed information for virus assembly must reside in the protein alone.

Ultraviolet absorption and ultraviolet inactivation studies on DNA in Pf1 indicate that the DNA bases are stacked on each other, even though the average separation between bases along each DNA strand is about 0.7 nm, twice the stacking separation (Wiseman *et al.* 1975). These results can be reconciled if bases on each strand are intercalated between bases on the opposite strand. Preliminary molecular model-building studies using the conformation calculated by de Coen & Kerckx (1970) suggest that such a self-intercalated DNA structure may be stereochemically feasible. Attempts to identify a DNA-binding protein like 5p in

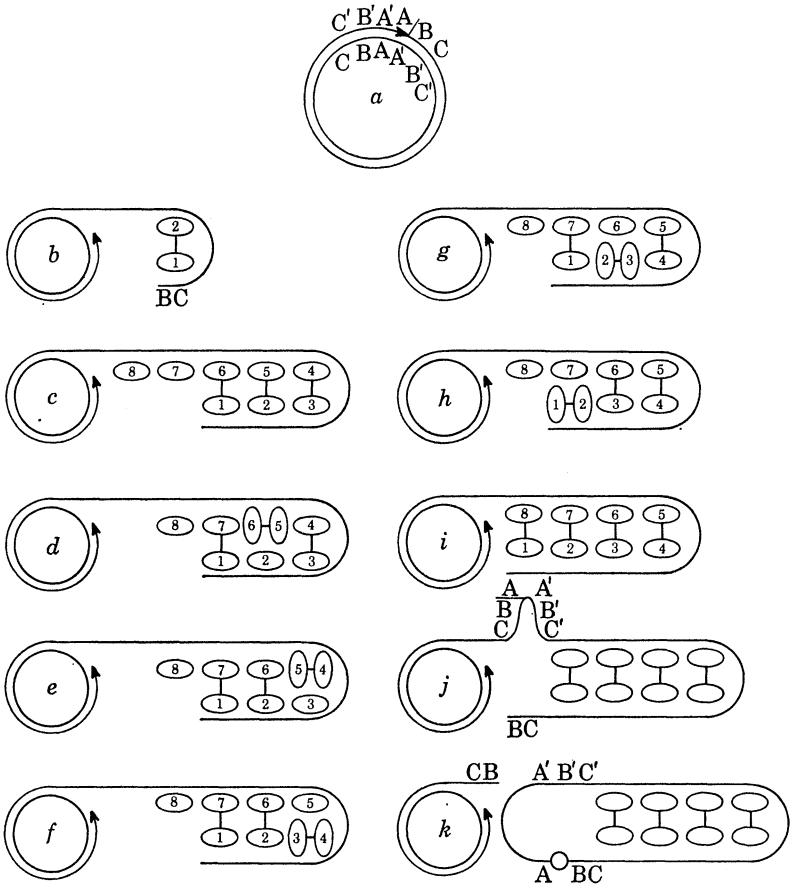


FIGURE 7. Model for assembly of the DNA-5p complex. (*a*) Circular duplex DNA (form I) is nicked at a specific site (Tabak *et al.* 1974) in a hairpin region with twofold sequence symmetry (Schaller *et al.* 1969). The dashed letters represent bases that pair with the corresponding undashed letters. (*b*) The single-strand DNA is replicated by a rolling circle mechanism. As the 5′ end of the DNA is displaced by the growing 3′ end, it forms a complex with a 5p dimer, causing the DNA to fold back on itself in a hairpin. (*c*)–(*i*) [The circular template is not to scale in this series.] As the hairpin grows by cooperative addition of 5p (Dunker 1975), the 5p molecules associate with newly revealed single-stranded regions, preventing synthesis of a complementary strand. The two strands are held together by dimerization between 5p molecules. They advance with respect to one another by propagation of a discontinuity. One type of possible propagation is illustrated. Molecules of 5p dimerize with adjacent partners on the same strand, at the expense of partners across strands, perhaps as a result of a flip-flop configurational change (Lazdunski 1972). (*j*) The advance of the growing DNA-5p complex is stopped after a unit length by a hairpin formed in the single-stranded DNA at the region of two-fold sequence symmetry. The hairpin is nicked by the same site-specific endonuclease that nicked the original form I. (*k*) The two ends of the newly made single-stranded DNA associate by complementary base-pairing, and the gap is sealed by a host ligase (Bouvier & Zinder 1974; A. Kornberg 1974).

bacteria infected with Pf1 have been unsuccessful (E. A. Anderson & D. A. Marvin, unpublished results). Perhaps 5p is necessary to hold the two strands of fd or Xf DNA in a pseudo-base-paired structure preparatory to insertion in the coat-protein tube, but self-intercalation alone is sufficient to hold the two strands of Pf1 DNA together.

Model for assembly

A reasonable hypothesis for the assembly of fd virions is illustrated in figure 8. The α-helical coat protein is inserted across the plasma membrane of the cell, with the *N*-terminus oriented outwards. The *N*-terminal peptide on nascent coat protein facilitates insertion into the membrane; cleavage of this peptide facilitates assembly of coat into virions.

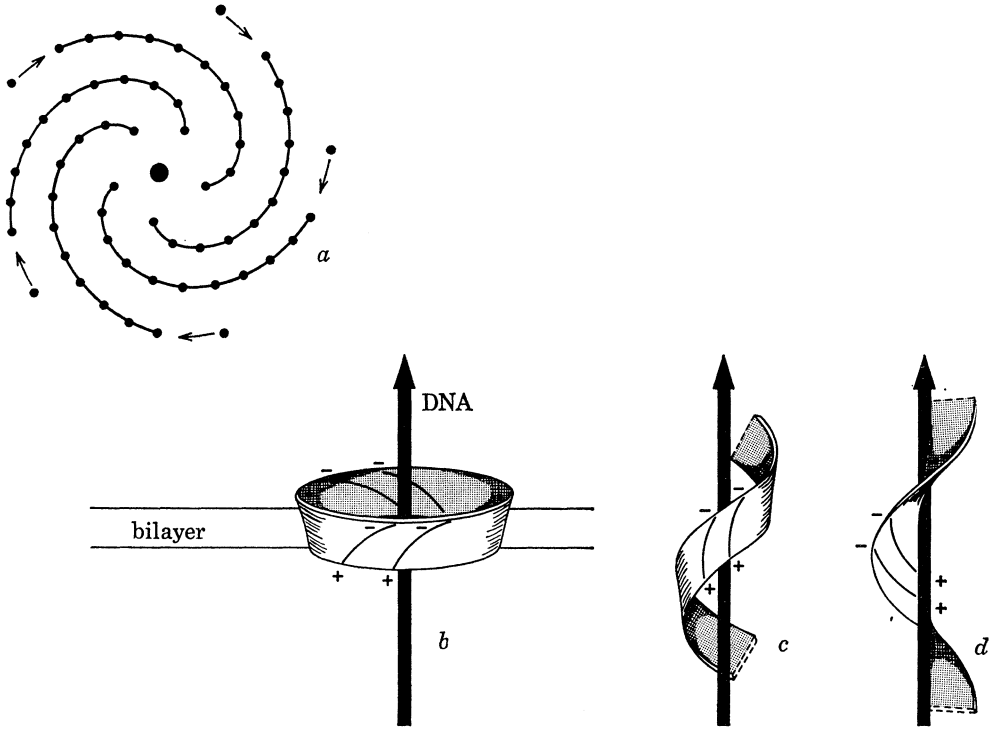


FIGURE 8. Model for assembly of the virion at the membrane. (*a*) Top view of proteins in the membrane. Large central dot is DNA in section (the two DNA strands are not distinguished); small dots are coat proteins in section. Coat protein molecules assemble through hydrophobic bonds into five interleaved spirals, that continually grow from the outer ends and are continually diminished by extrusion through the membrane from the inner end. Each spiral surface is considered as a unit held together by the 0–5 bonds, that slides from the membrane onto the virion. (*b–d*) A detailed suggestion of the way in which one of these five spirals assembles into one of the 0–5 helicoids, based on a model of the bending of a catenoid into a helicoid. (*b*) The coat protein molecules surround the DNA, with the positive *C*-terminus nearer the DNA. The spiral surface is here simplified to a catenoid. (*c–d*) The catenoid can be bent into a helicoid with no local deformation of the surface, and therefore no change in the nature of the bonds between coat protein molecules within the surface (Struik 1961).

While the coat protein is being placed in the membrane, DNA-5p complexes are formed as suggested in figure 7. Initiation of extrusion of the virion may require the minor coat protein as a pilot protein to lead the DNA across the membrane, as found for penetration of the DNA upon infection (A. Kornberg 1974), although apparently the minor protein does not require a specific attachment sequence on the DNA (Knippers & Hoffmann-Berling 1966; Tate & Petersen 1974). Once extrusion has started, the 5p is displaced from the DNA by the high

local charge on the inner surface of the lipid bilayer [DNA-5p complexes are disrupted by ionic strength above 0.1 M (Pratt *et al.* 1974)]. The DNA passes into the membrane, through a pore-like assembly of coat protein molecules. Here it forms a condensation centre for the coat proteins as it is continually extruded.

Although assembly and extrusion of the virion does not lyse or kill the cell, failure of the assembly process does kill the cell. This killing involves viral DNA replication (Timmis & Marvin 1974*a*), the gene 5 protein (Timmis & Marvin 1974*b*) and the major coat protein (Webster & Cashman 1973), and appears to act by disrupting cell respiration (Hohn, von Schütz & Marvin 1971). Perhaps the α-helical coat proteins spanning the membrane can form pores for protons, as in purple membrane of *Halobacterium halobium* (Henderson & Unwin 1975). If viral assembly proceeds normally, the α-helices would be unavailable for protons, but if assembly is stopped, the α-helices would become available to permit leakage of protons, thereby destroying the proton gradient required for respiration (Mitchell 1972).

Discussion

The structure of macromolecular assemblies cannot usually be determined by simple crystallographic methods. Three-dimensional diffraction data is seldom available; when it is, it seldom extends to atomic resolution. Additional facts are often used to support the structure analysis. One simple but powerful fact about molecular assemblies is that they are built out of subunits. Just as a small molecule can be considered as an arrangement of atoms, so may a large assembly be considered as an arrangement of subunits. The structure analysis can then be divided into two stages: first determine the structure of the subunits, and then determine the arrangement of the subunits in the assembly. Regular viruses, since they contain many identical subunits, are especially convenient examples.

In the case of filamentous bacterial viruses, there is good evidence that the subunit can be approximated by a single length of α-helix: a known structure. The evidence includes not only the early indications from the diffraction pattern and from spectroscopic measurements, but also those predictions made by the detailed model that are now confirmed by fact. The nearest neighbour interactions between subunits are those predicted by the theory of packing of elongated subunits; the proteins when in the membrane extend across the lipid bilayer as expected for an elongated subunit; the model predicts the observed diffraction data. The convergence of the calculated Fourier transform to the observed is an especially important condition for confidence in any model. We have now begun be to use energy minimization techniques (Levitt & Lifson 1969) to optimize the stereochemistry of the model (figure 3). Comparative studies between various class I strains, or between class I and class II, will give more confidence in the model than study of one strain of virion alone. Further study will require more accurate diffraction data to refine the model using difference Fourier and/or least squares techniques. Radial density calculations for virions having heavy atoms attached to specific residues could be used to define the radius of the heavy atom and thereby the radius of the residue. Difference Fourier studies between intact virions and protein shells lacking DNA could be used to define the DNA conformation more precisely. Chemical modification studies to define externally accessible amino acid residues, and chemical crosslinking studies (protein–protein or protein–DNA) to define adjacent residues would place further constraints on the model. Because of the small size of the subunit, these approaches should yield the structure of filamentous bacterial viruses at atomic resolution.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Knowledge of static structure is interesting not only in itself, but also for the insights that it can give into dynamic function. Experiments from other laboratories on the DNA-5p complex and on the location of coat proteins in the membrane, together with our results on the structure of the virion, have suggested new mechanisms of molecular assembly. The displacement of 5p by coat protein must involve some dynamic change in DNA structure, indicating that the 5p is more than just a scaffold for assembly. The emergence of the coat protein from the membrane onto the DNA must also involve dynamic DNA–protein–lipid interactions. To advance, it will be desirable to know the structure of the DNA-5p complex. This could be accomplished by determining the structure of the 5p subunit using single-crystal methods, followed by fibre-diffraction studies of the DNA-5p complex as a whole to determine the arrangement of subunits in the assembly. The arrangement of coat protein molecules in the lipid bilayer could be studied by reconstructing protein–bilayer assemblies *in vitro*. Knowledge of the molecular structure of these precursors in assembly will permit formulation of more detailed hypotheses about mechanisms of assembly.

References (Marvin & Wachtel)

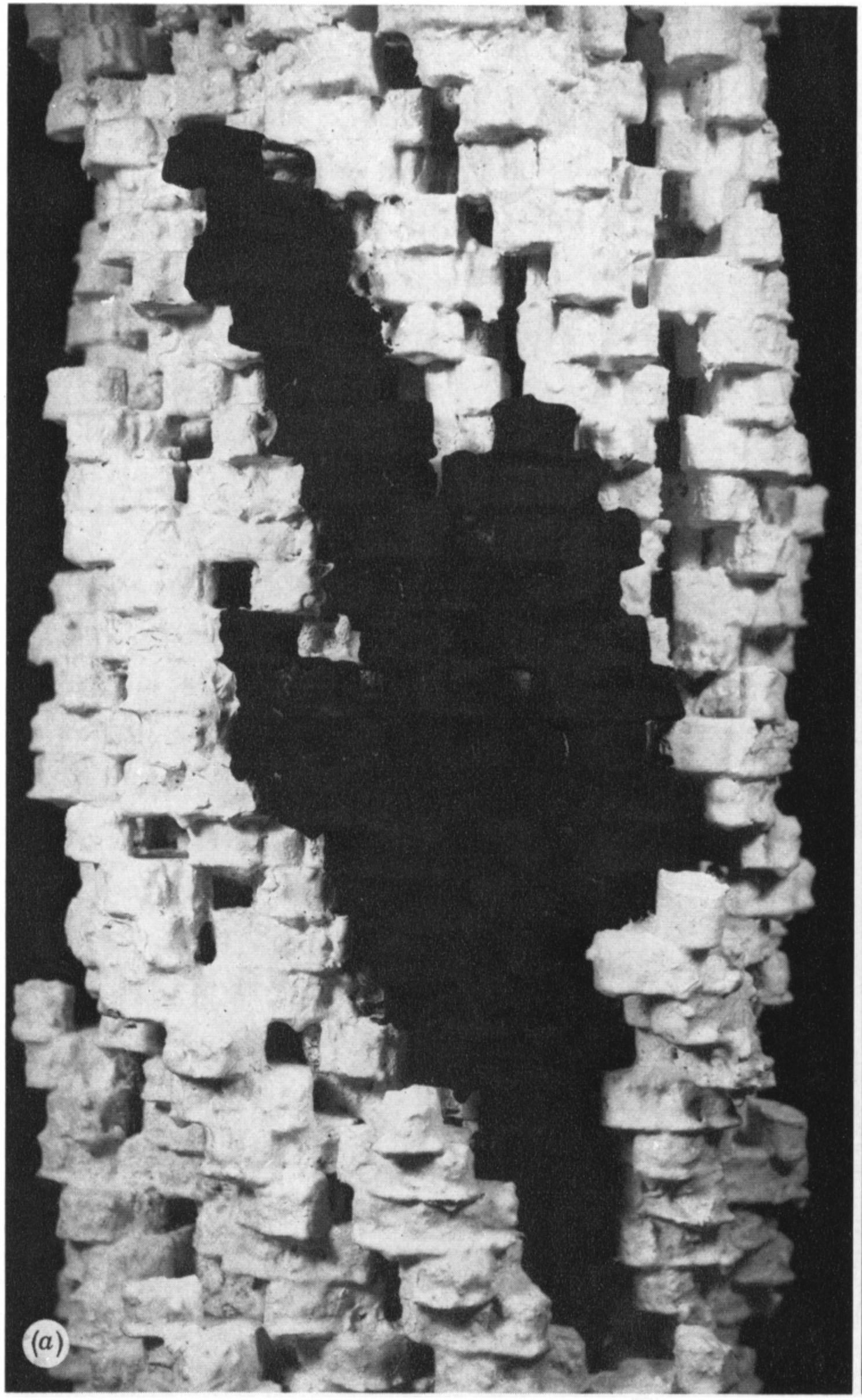
Adler, I. 1974 *J. theor. Biol.* **45**, 1–79.
Alberts, B., Frey, L. & Delius, H. 1972 *J. molec. Biol.* **68**, 139–152.
Anderson, R. A., Nakashima, Y. & Coleman, J. E. 1975 *Biochemistry* **14**, 907–917.
Blaurock, A. E. 1975 *J. molec. Biol.* **93**, 139–158.
de Boer, J. H. 1968 *The dynamical character of adsorption*, 2nd edn. Oxford University Press.
Bouvier, F. & Zinder, N. D. 1974 *Virology* **60**, 139–150.
Chothia, C. 1975 *Nature, Lond.* **254**, 304–308.
de Coen, J. L. & Kerckx, J. 1970 *FEBS Lett.* **11**, 241–245.
Cohen, C. & Holmes, K. C. 1963 *J. molec. Biol.* **6**, 423–432.
Coxeter, H. S. M. 1969 *Introduction to geometry*. New York: Wiley.
Crick, F. H. C. 1953*a* *Acta Crystallogr.* **6**, 685–689.
Crick, F. H. C. 1953*b* *Acta Crystallogr.* **6**, 689–697.
Day, L. A. 1966 *J. molec. Biol.* **15**, 395–398.
Day, L. A. 1969 *J. molec. Biol.* **39**, 265–277.
Day, L. A. 1973 *Biochemistry* **12**, 5329–5339.
Diamond, R. 1966 *Acta Crystallogr.* **21**, 253–266.
Dunker, A. K. 1975 *Biochim. biophys. Acta*, **402**, 31–35.
Eiserling, F. A. & Dickson, R. C. 1972 *A. Rev. Biochem.* **41**, 467–502.
Fedorov, B. A., Ptitsyn, O. B. & Voronin, L. A. 1974 *J. appl. Cryst.* **7**, 181–186.
Fraser, R. D. B., MacRae, T. P. & Miller, A. 1965 *J. molec. Biol.* **14**, 432–442.
Geider, K. & Kornberg, A. 1974 *J. biol. Chem.* **249**, 3999–4005.
Green, N. M. & Flanagan, M. T. 1976 *Biochem. J.* **153**, 729–732.
Henderson, R. & Unwin, P. N. T. 1975 *Nature, Lond.* **257**, 28–32.
Hoffmann-Berling, H. & Mazé, R. 1964 *Virology* **22**, 305–313.
Hohn, B., von Schütz, H. & Marvin, D. A. 1971 *J. molec. Biol.* **56**, 155–165.
Klug, A., Crick, F. H. C. & Wyckoff, H. W. 1958 *Acta Crystallogr.* **11**, 199–213.
Knippers, R. & Hoffmann-Berling, H. 1966 *J. molec. Biol.* **21**, 305–312.
Konings, R. N. H., Hulsebos, T. & van den Hondel, C. A. 1975 *J. Virol.* **15**, 570–584.
Kornberg, A. 1974 *DNA replication*. San Francisco: W. H. Freeman.
Kornberg, R. D. 1974 *Science, N.Y.* **184**, 868–871.
Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. 1960*a* *J. molec. Biol.* **2**, 38–64.
Langridge, R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. 1960*b* *J. molec. Biol.* **2**, 19–37.
Lazdunski, M. 1972 In *Current topics in cellular regulation* (ed. B. L. Horecker & E. R. Stadtman), p. 267. New York: Academic Press.
Lee, B. & Richards, F. M. 1971 *J. molec. Biol.* **55**, 379–400.
Levitt, M. & Lifson, S. 1969 *J. molec. Biol.* **46**, 269–279.
Levitt, M. & Warshel, A. 1975 *Nature, Lond.* **253**, 694–698.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Lin, N. S.-C. & Pratt, D. 1974 *Virology* **61**, 334–342.
Lipson, H. & Cochran, W. 1953 *The determination of crystal structures*. London: G. Bell and Sons, Ltd.
Marvin, D. A. 1966 *J. molec. Biol.* **15**, 8–17.
Marvin, D. A. & Hohn, B. 1969 *Bact. Rev.* **33**, 172–209.
Marvin, D. A., Pigram, W. J., Wiseman, R. L., Wachtel, E. J. & Marvin, F. J. 1974*a* *J. molec. Biol.* **88**, 581–600.
Marvin, D. A., Wiseman, R. L. & Wachtel, E. J. 1974*b* *J. molec. Biol.* **82**, 121–138.
Marvin, D. A. & Wachtel, E. J. 1975 *Nature, Lond.* **253**, 19–23.
Mazur, B. J. & Model, P. 1973 *J. molec. Biol.* **78**, 285–300.
Mitchell, P. 1972 *Bioenergetics* **3**, 5–24.
Nakashima, Y., Dunker, A. K., Marvin, D. A. & Konigsberg, W. 1974 *FEBS Lett.* **40**, 290–292; **43**, 125.
Nakashima, Y., Wiseman, R. L., Konigsberg, W. & Marvin, D. A. 1975 *Nature, Lond.* **253**, 68–71.
Ninio, J., Luzzati, V. & Yaniv, M. 1972 *J. molec. Biol.* **71**, 217–229.
Oey, J. L. & Knippers, R. 1972 *J. molec. Biol.* **68**, 125–138.
Pratt, D., Laws, P. & Griffith, J. 1974 *J. molec. Biol.* **82**, 425–439.
Ramachandran, G. N. & Sasisekharan, V. 1968 *Adv. Protein Chem.* **23**, 283–437.
Rasched, I. & Pohl, F. M. 1974 *FEBS Lett.* **46**, 115–118.
Salstrom, J. S. & Pratt, D. 1971 *J. molec. Biol.* **61**, 489–501.
Schaller, H., Voss, H. & Gucker, S. 1969 *J. molec. Biol.* **44**, 445–458.
Smilowitz, H., Carson, J. & Robbins, P. W. 1972 *J. supramol. Struct.* **1**, 8–18.
Squire, J. 1973 *J. molec. Biol.* **77**, 291–323.
Staudenbauer, W. L. & Hofschneider, P. H. 1973 *Eur. J. Biochem.* **34**, 569–576.
Struik, D. J. 1961 *Differential geometry*, 2nd edn. Addison-Wesley Publishing Co. p. 121.
Tabak, H. F., Griffith, J., Geider, K., Schaller, H. & Kornberg, A. 1974 *J. biol. Chem.* **249**, 3049–3054.
Tate, W. P. & Petersen, G. B. 1974 *Virology* **62**, 17–25.
Thomas, G. J. & Murphy, P. 1975 *Science, N.Y.* **188**, 1205–1207.
Timmis, K. & Marvin, D. A. 1974*a* *Virology* **59**, 281–292.
Timmis, K. & Marvin, D. A. 1974*b* *Virology* **59**, 293–300.
Trenkner, E. 1970 *Virology* **40**, 18–22.
Wachtel, E. J., Wiseman, R. L., Pigram, W. J., Marvin, D. A. & Manuelidis, L. 1974 *J. molec. Biol.* **88**, 601–618.
Webster, R. E. & Cashman, J. S. 1973 *Virology* **55**, 20–38.
Wiseman, R. L., Berkowitz, S. A. & Day, L. A. 1975 *Fedn Proc.* **34**, 637.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



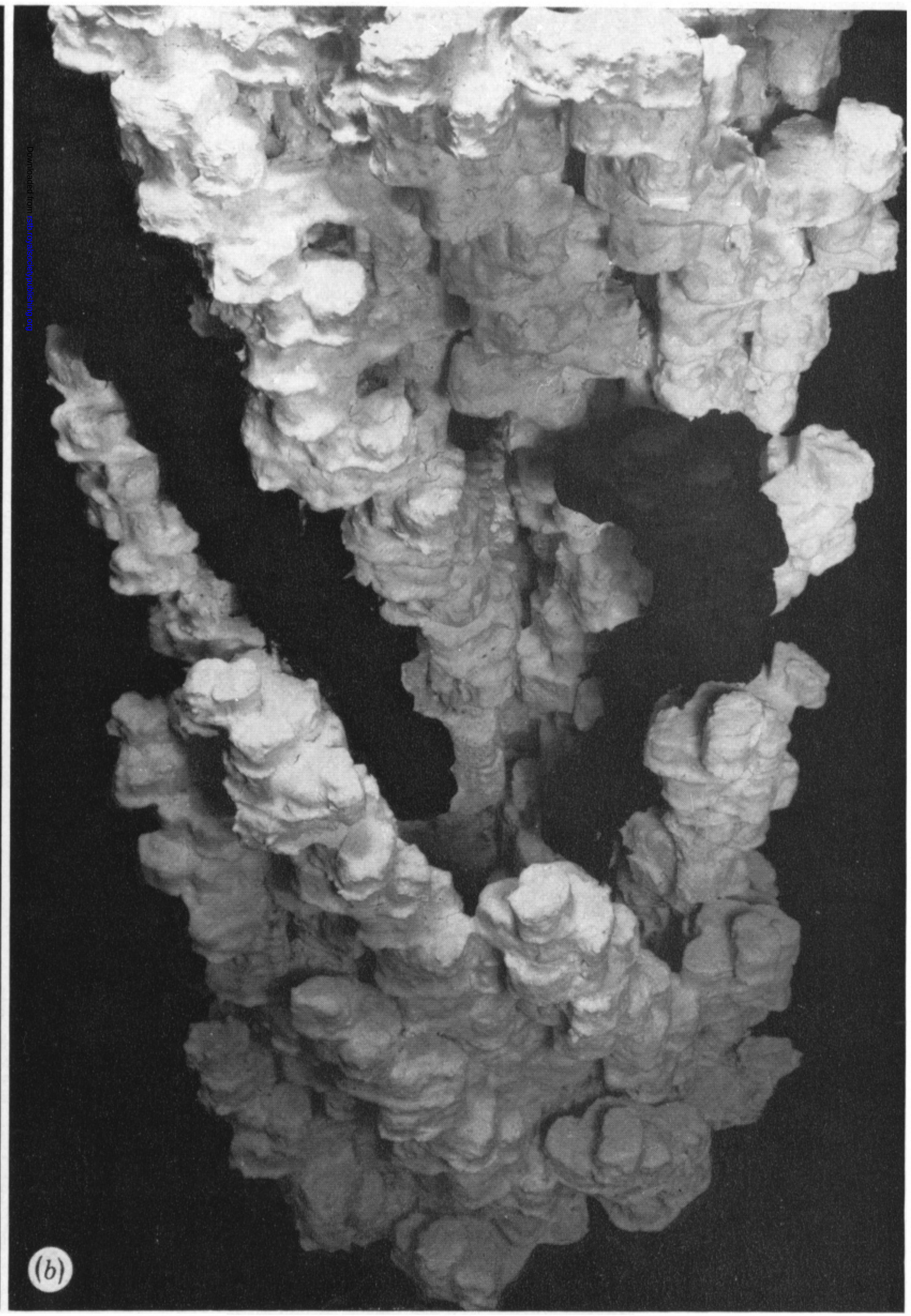


FIGURE 5. For description see opposite.