J. Mol. Biol. (1994) **235**, 260–286

Molecular Models and Structural Comparisons of Native and Mutant Class I Filamentous Bacteriophages

Ff (fd, f1, M13), If1 and IKe

D. A. Marvin†, R. D. Hale‡

Department of Biochemistry, University of Cambridge Cambridge CB2 1QW, U.K.

C. Nave

SERC Daresbury Laboratory Warrington WA4 4AD, U.K.

and M. Helmer Citterich

Dpto di Biologia, II Universita di Roma, Tor Vergata, 00173 Roma, Italy

The filamentous bacteriophages are flexible rods about 1 to 2 μm long and 6 nm in diameter, with a helical shell of protein subunits surrounding a DNA core. The approximately 50-residue coat protein subunit is largely α-helix and the axis of the α-helix makes a small angle with the axis of the virion. The protein shell can be considered in three sections: the outer surface, occupied by the N-terminal region of the subunit, rich in acidic residues that interact with the surrounding solvent and give the virion a low isoelectric point; the interior of the shell, including a 19-residue stretch of apolar side-chains, where protein subunits interact mainly with each other; and the inner surface, occupied by the C-terminal region of the subunit, rich in basic residues that interact with the DNA core. The fact that virtually all protein side-chain interactions are between different subunits in the coat protein array, rather than within subunits, makes this a useful model system for studies of interactions between α-helix subunits in a macromolecular assembly.

We describe molecular models of the class I filamentous bacteriophages. This class includes strains fd, f1, M13 (these 3 very similar strains are members of the Ff group), If1 and IKe. Our model of fd has been refined to fit quantitative X-ray fibre diffraction data to 30 Å resolution in the meridional direction and 7 Å resolution in the equatorial direction. A simulated 3·3 Å resolution diffraction pattern from this model has the same general distribution of intensity as the experimental diffraction pattern. The observed diffraction data at 7 Å resolution are fitted much better by the calculated diffraction pattern of our molecular model than by that of a model in which the α-helix subunit is represented by a rod of uniform density.

The fact that our fd model explains the fd diffraction data is only part of our structure analysis. The atomic details of the model are supported by non-diffraction data, in part previously published and in part newly reported here. These data include information about permitted or forbidden side-chain replacements, about the effect of chemical modification, and about spectroscopic experiments. The side-chain comparisons include the other class I wild-type strains, which have similar diffraction patterns and similar overall distributions of amino-acid residues by type, but different detailed sequences; previously reported mutants of Ff strains, especially in the acidic N-terminal region and the basic C-terminal region; and new mutants in the interior of the protein shell, some of which are viable (F11Y, Y21M, I22V and Y24M), and others of which carry apparently lethal mutations (F11M, A27P and exchanges of Ile22 and Ile32 for other apolar residues).

Keywords: α-helix; fibre diffraction; *Inovirus*; macromolecular assembly; virus structure

† Author to whom all correspondence should be addressed.

‡ Present address: ABC Ltd, 307 Huntingdon Road, Cambridge CB3 0JX, U.K.

0022-2836/94/010260-27 $08.00/0

1. Introduction

Filamentous bacteriophages (*Inovirus*) are assembled as they extrude through the bacterial membrane without lysing or otherwise killing the bacterial host. The major coat protein (g8p†) is inserted into the bacterial inner membrane as a membrane-spanning precursor protein with a short leader sequence which is then removed. The single-stranded circular viral DNA is passed from an intracellular complex with the g5p DNA-binding protein to the final assembly with the g8p. The virion (the virus particle) of all *Inovirus* strains is a flexible nucleoprotein rod about 6 nm in diameter and 1 to 2 μm long, comprising a helical tube of g8p subunits surrounding a core of DNA. A few minor proteins that are involved in initiating (g7p and g9p) and terminating (g3p and g6p) the assembly process are found at the two ends of the completed virion. Foreign DNA can be inserted into the genome and carried through future generations, either as a separately expressed gene or as an addition to an existing gene. Understanding the molecular structure of the virion is essential to understand in detail this complex and interesting process of macromolecular assembly (for general reviews, see Rasched & Oberer, 1986; Model & Russel, 1988; for a review of the virion assembly process, see Russel, 1991).

Virions from different *Inovirus* strains have similar molecular architecture but different coat protein sequences and different values of some structural parameters. The g8p is largely α-helix, with about 50 amino-acid residues ($\pm 10\%$). There are two symmetry classes of *Inovirus*. Class I includes strains fd, If1 and IKe. Class II includes strains Pf1 and Xf. Strain fd belongs to the Ff group within class I. The members of the Ff group have virtually identical g8p sequences and only slightly different genome sequences. Two other commonly studied Ff strains are f1, which has the same g8p sequence as fd, and M13, which has an Asp to Asn exchange at position 12. The length of the virion depends on the size of the DNA and on the positive charge density along the inside surface of the protein tube. The wild-type DNA has about 6500 nucleotides ($\pm 10\%$), but longer DNA can be encapsidated by simply adding more subunits to the protein tube during assembly (for a review of the virion structure, see Marvin, 1990).

X-ray fibre diffraction patterns of the virion show that the long axis of the α-helical protein subunit lies at a small angle to the long axis of the virion. Direct interpretation of the strong intensity distribution on low-resolution diffraction patterns, followed by molecular model-building, energy refinement, and comparison of calculated Fourier transforms with observed diffraction data, gave the orientation of the α-helix within the surface lattice of the virion and indicated a common subunit shape and a common molecular architecture for all strains (Marvin *et al.*, 1974*a*, *b*; Marvin & Wachtel, 1975, 1976). The proposal of a common architecture was supported by the discovery that, despite differences in detail, both class I and class II g8p sequences have the same characteristic distribution of amino-acid residues, with several acidic residues near the N terminus, a 19-residue apolar domain near the middle and a cluster of basic residues near the C terminus (Nakashima *et al.*, 1975). The subunits in the model are gently curved α-helix rods in an overlapping interdigitated helical array. Each subunit slopes from the outer to the inner radius of the virion, with the acidic N terminus facing outwards to explain the low isoelectric point of the virion, the apolar region maintaining interactions between subunits, as indicated by the sensitivity of the assembly to organic solvents and its resistance to extremes of pH and ionic strength, and the basic residues facing inwards to neutralize the charge on the DNA at the core. The neighbours of an arbitrary subunit, indexed 0, have indices along the helical array defined by the Fibonacci sequence 1, q, $1 + q$, $1 + 2q$, $2 + 3q$ (with q a positive integer), as discussed in detail by Marvin (1989, 1990).

The early X-ray data could not resolve some symmetry ambiguities: the choice of enantiomorph (the hand of the helix followed by the subunit axis); the virion helix symmetry (is q 4 or 5?); and the relationship between the class I and class II symmetries. Although the ambiguities of symmetry do not affect conclusions about the architecture of the virion, it was necessary to resolve them to be able to proceed with more detailed structure analysis. Several attempts, in part successful, were made to infer the symmetry (Marvin & Wachtel, 1976; Day & Wiseman, 1978; Makowski & Caspar, 1978, 1981), but none gave the correct enantiomorph. The development of methods using the diamagnetic anisotropy of *Inovirus* to improve alignment in fibres (Nave *et al.*, 1979; Torbet & Maret, 1979; Maret & Dransfeld, 1985) enabled closely spaced layer lines to be resolved and small intensity differences between native and heavy-atom derivative

† Abbreviations used: To describe gene products, we use abbreviations of the form g8p for the gene *8* protein (in its processed form with the leader sequence removed, if appropriate).

To describe mutants of the virion, we use the notation "virus strain/gene number/original residue type/residue number/new residue type". For example, fdg8K48Q is a mutant of fd in which K at position 48 of g8p has been replaced by Q; this may be shortened to simply K48Q if no ambiguity would be created.

We compare aligned amino acid sequences of fd, If1 and IKe using square brackets to define position number, which is not the same as the distance along the If1 or IKe sequence (see Table 1). For example Lys[48] (or Lys48) of fd corresponds to Arg[48] (or Arg51) of IKe.

To discuss interactions between an arbitrary origin subunit $k = 0$ and a neighbouring subunit k, we use the notation "subunit index/residue type/residue number". For example residue 0Trp26 is residue Trp26 on subunit $k = 0$; residue 11Tyr21 is residue Tyr21 on subunit $k = 11$. If no ambiguity would be created, the index 0 may be omitted.

virions to be measured. These experimental advances enabled resolution of the symmetry ambiguities (Banner *et al.*, 1981; Nave *et al.*, 1981; Bryan *et al.*, 1983; Bryan, 1987; Marvin *et al.*, 1987; Glucksman *et al.*, 1992). The subunit axis follows a right-handed helix; the virion helix symmetry obeys $q = 5$; and class I can be described as a step-function perturbation of the class II helix.

Under standard external conditions, the three class I strains fd, If1 and IKe have similar virion helix symmetry and similar high α-helix content in the subunit despite substantial differences in the g8p sequence (Table 1). Marvin (1990) defined this common symmetry as the "canonical" symmetry, having a 1·97 *u/t* screw axis along the virion axis relating successive groups of five g8p subunits which are themselves related by a 5-fold rotation around the virion axis. Molecular models of the protein coat of fd, If1 and IKe in the canonical 1·97 *u/t* symmetry were described by Marvin (1990), and the co-ordinates of the fd model have been deposited in the Protein Data Bank (Bernstein *et al.*, 1977) under entry 1IFD. Here we discuss and compare molecular models of wild-type and mutant class I virions in the canonical symmetry. The fd symmetry can be changed slightly by external agents (such as pH) from the canonical 1·97 *u/t* symmetry (fd^C) to a 2·00 *u/t* diad symmetry (fd^D), and we also discuss models with the fd^D symmetry to illustrate the small changes in the subunit that accompany changes in the symmetry.

2. Materials and Methods

(a) *Experimental methods*

Mutants of f1 were generated by site-directed mutagenesis using standard methods (Sambrook *et al.*, 1989). The 5 oligonucleotides used to prepare the f1 mutants were:

R52: Ile or Val or Leu or Phe in place of Ile22

5′-ACGCATAACCAANATATTCGGTC-3′

R53: Ile or Val or Leu or Phe in place of Ile32

5′-TTGCGCCGACGANGACAACAACC-3′

R60: Tyr21 to Met

5′-CATAACCGATCATTTCGGTCGCT-3′

R61: Ala27 to Pro

5′-ACAACCATCGGCCACGCATAA-3′

R104: Tyr24 to Met

5′-CATCGCCCACGCCATACCGATATATTC-3′

The R52 and R53 oligonucleotides should produce each of the 4 replacements at the same frequency. Although the Ile replacement gives a wild-type protein, at the DNA level it is a mutant. This provides an internal control for mutagenesis frequency. The mutagenesis method was that of Kunkel as described by Sambrook *et al.* (1989), involving selection against template strands containing uracil.

Wild-type and mutant virions for X-ray diffraction experiments were grown and purified using standard methods (e.g. Nave *et al.*, 1981) and purified samples were dialysed into 10 mM Tris·HCl, 1 mM EDTA (pH 8·0). Fibre preparation and X-ray diffraction techniques were essentially as described for Pf1 (Nave *et al.*, 1981).

(b) *Calculations*

The methods that we use for X-ray data processing, Fourier transform calculation and model refinement of *Inovirus* have been described (Nave *et al.*, 1981; Marvin & Nave, 1982; Bryan, 1987; Marvin *et al.*, 1987, 1992; Marvin, 1990). Intensity is confined to a set of layer lines of index $l = Zc$, where Z is the reciprocal space distance normal to the layer line and c is the helix repeat. For a single Bessel function term of order n, the amplitude of the Fourier-Bessel transform as a function of distance R along a layer line l is $G_{nl}(R)$. If there is an N-fold rotation axis relating asymmetric units, the Bessel function terms are systematically absent unless n is an integral multiple of N. Diffracted intensity on layer line l is $I_l(R) = \Sigma_n |G_{nl}(R)|^2$. In the low resolution region (spacings greater than about 10 Å), calculated diffraction patterns of biological macromolecules cannot fit the observed data unless solvent is included. In single crystal studies the space between the molecules can be filled with solvent, either as explicit solvent molecules or as uniform electron density. In our fibre diffraction studies, there is no well-defined crystal unit cell and we include solvent by defining all electron density relative to uniform solvent density. Then outside the molecular boundary (Richards, 1985) the relative electron density is zero, and within the molecular boundary the relative density is that of the molecule minus that of uniform solvent filling the molecular boundary.

Fibre diffraction data seldom extend either to high resolution or to 3 dimensions, and such diffraction data cannot uniquely define atomic positions in a molecular model. Nevertheless we use molecular models right from the start of our structure analysis. Non-bonded interchain contacts between nearest-neighbour asymmetric units in the assembly are as close as intrachain contacts in many globular proteins, and provide important constraints on the model; models also enable testing against non-diffraction information. Since the diffraction data are so limited, we reduce the effective number of independent variables in the model by constraining backbone (ϕ, ψ) torsion angles and O_i to N_{i+4} hydrogen-bond lengths to be near the values found for α-helices (Baker & Hubbard, 1984; Morris *et al.*, 1992). We also define side-chain torsion angles at the start of the refinement to have the idealized values for α-helices. Side-chains in proteins solved at high resolution are found to be limited to a relatively small set of conformations, and this limitation is especially stringent for χ^1 of side-chains on α-helices (McGregor *et al.*, 1987; Ponder & Richards, 1987; Summers *et al.*, 1987). For many side-chain conformations surveyed in proteins, and especially for those in α-helix regions, χ^1 is in the $-60°$ rotamer (that is, $-120° < \chi^1 < 0°$). For Val, the 180° rotamer is more frequent, but this is a trivial consequence of the IUPAC-IUB conventions, and here we use χ^1 for Val measured to the $C^{\gamma 2}$, as discussed by Summers *et al.* (1987). We supplement the diffraction data and the stereochemical data with specific chemical and spectroscopic data that define the absolute or relative positions of specific side-chains in the model, and genetic data that define the effect on the structure of specific side-chain exchanges.

We refine models using the Levitt energy refinement program (Levitt, 1974, 1983), modified to include van der Waals contacts between neighbouring symmetry-related subunits. We take convergence of refinement as the point

Table 1

Class I coat protein sequences and environment

	Sequence			Environment‡		
Position†	fd	If1	IKe	fd	If1	IKe
−2			A			E
−1			E−			P1
0		A	P		P1	B1
1	A	D−	N	E	P2	P2
2	E−	D−	A	E	E	E
3	G	A	A	E	P1	P1
4	D−	T	T	P1	P1	P1
5	D−	S	N	E	E	E
6	P	Q	Y	P2	P2	P2
7	A	A	A	P1	P1	P1
8	K+	K+	T	B3	P2	P2
9	A	A	E−	E	E	E
10	A	A	A	P1	P1	P1
11	F	F	M	B1	B1	B1
12	D−	D−	D−	P2	P2	P2
13	S	S	S	E	P2	P2
14	L	L	L	B2	B1	B1
15	Q	T	K+	B2	P1	B3
16	A	A	T	E	E	E
17	S	Q	Q	P2	B3	B3
18	A	A	A	P1	P1	P1
19	T	T	I	P2	P1	P1
20	E−	E−	D−	P2	P2	P2
21	Y	M	L	B1	B1	B2
22	I	S	I	B1	P1	B1
23	G	G	S	E	E	P1
24	Y	Y	Q	B3	B3	B3
25	A	A	T	P1	P1	P1
26	W	W	W	B2	B2	B2
27	A	A	P	E	E	P2
28	M	L	V	B1	B1	B1
29	V	V	V	B1	B1	B1
30	V	V	T	P1	P1	P1
31	V	L	T	P2	P2	P1
32	I	V	V	B1	B1	B1
33	V	V	V	B2	B1	B1
34	G	G	V	E	E	B1
35	A	A	A	P1	P1	P1
36	T	T	G	P1	P1	E
37	I	V	L	B1	B1	B1
38	G	G	V	E	E	B2
39	I	I	I	B1	B1	B1
40	K+	K+	R+	B3	B3	B3
41	L	L	L	B1	B1	B1
42	F	F	F	B1	B1	B1
43	K+	K+	K+	B3	B3	B3
44	K+	K+	K+	B3	B3	B2
45	F	F	F	B1	B1	B1
46	T	V	S	P2	B3	P1
47	S	S	S	E	E	E
48	K+	R+	R+	B1	B1	B2
49	A	A	A	P1	P1	P1
50	S	S	V	P2	P2	P2

†Position in the set of aligned sequences is defined by the fd sequence, so sequences that extend beyond the N terminus of fd are given negative position numbers. These negative position numbers should not be confused with the negative numbers assigned by some authors to positions in the gene *8* leader sequence. The sequence of the Ff strain f1 is identical to that of fd, and M13 differs from that of fd by a single Asp to Asn exchange at position 12. Positively charged residues are marked + and negatively charged residues are marked −. Sequences are taken from Model & Russel (1988), who give the original literature references.

‡Environment categories for side-chains on a single subunit surrounded by its nearest neighbours in the canonical assembly. The environment categories are defined in detail by Bowie *et al.* (1991). Abbreviations are: B, buried; P, partially buried; E, exposed, as measured by total side-chain surface area covered by other atoms in the structure. Increasing value of the numbers on categories B and P indicate an increasing fraction of side-chain surface area covered either by solvent or by polar (N and O) atoms in other parts of the assembly. The secondary structure was assumed to be all α-helix.

where the standard deviation of C–C bond lengths from the ideal value is less than 0·03 Å, corresponding to an energy about 0·2 kcal/mol above the minimum energy for a single C–C bond. Deviation from the ideal of other bond lengths, bond angles, torsion angles, non-planar distortions and van der Waals contacts are within this energy range.

We use the method of Jack & Levitt (1978) to minimize the difference between calculated and observed diffraction for quantitative X-ray data by way of a difference electron density map calculated within the molecular volume of a single asymmetric unit of the virion helix with respect to solvent background. We do not refine the temperature factor B, but instead use a constant $B = 10$ Å^2. Where Bessel function terms for several (n, l) overlap on the same layer line l, observed intensity $I_l(R)$ was separated between different n in the same ratio as the calculated intensities $|G_{nl}(R)|^2$. We call the observed intensity that has been separated in this way the "observed $|G_n(R)|^2$", and compare its square root, the "observed $|G_n(R)|$", with the $|G_n(R)|$ calculated from the model. To calculate the "observed" electron density distribution for the difference map, phases calculated for the current model were used with the observed amplitudes. The resultant map is biased towards the model, but iteration of this procedure minimizes the difference between observed and calculated $|G_n|$. The R-factor commonly used to assess single-crystal structure analyses is not ideal for assessing the fit between calculated and observed continuous transform fibre diffraction amplitudes, and we use the correlation coefficient to compare models. We determine the correlation coefficient between calculated and observed $|G_n|$ on layer-lines other than the equator, because the calculated diffraction on the equator is especially sensitive to the precise DNA model and the precise solvent density. For the work described here, this calculation is limited to $l = 1$, where it defines the lateral packing of the α-helices.

The arrangement of subunits in the surface lattice and the indexing system that we use to define the different subunits are illustrated by Marvin (1989, 1990). To generate co-ordinates $x_j(k)$, $y_j(k)$, $z_j(k)$ of the j^{th} atom in unit k from the given co-ordinates x_j, y_j, z_j of the j^{th} atom in unit 0 in a class I helix with parameters (τ, p), we write eqn (3) of Marvin (1990) as:

$$r_j(k) = r_j, \tag{1a}$$

$$\phi_j(k) = \phi_j + 72k + [k/5]\tau, \tag{1b}$$

$$z_j(k) = z_j + [k/5]p. \tag{1c}$$

where $[k/5]$ is the integral part of $k/5$, and angles are in degrees. The subunits can be considered in groups of 5 that are invariant to a 5-fold rotation around the z axis. One group of 5 subunits is related to the next along a helix of unit rise p and pitch P, so the unit twist relating one group to the next is p/P turn or $\tau = 360p/P$ degrees. The neighbours in contact with unit 0 are units $k = \pm 1$, ± 5, ± 6, ± 11 and ± 17. These symmetry-related copies were used to determine the energy of interchain van der Waals contacts during energy minimization. Contacts between the reference subunit and the k^{th} subunit are called 0 to k contacts. The 0 to k contacts are identical to the $(-k)$ to 0 contacts. The N-start helices are a set of N helices, each of which passes through subunits k, $k \pm N$, $k \pm 2N$ .... The path followed by the N-start helix from the 0 subunit is called the 0 to N direction. The canonical class I helix parameters are $\tau = -33{\cdot}23°$, $p = 16{\cdot}00$ Å in eqns (1a) to (1c). The co-ordinates of the asymmetric unit slewed into a new unit cell with slightly different parameters (τ', p') were determined using eqn (5) of Marvin (1990). For the perfect screw dyad symmetry, $\tau' = -36{\cdot}00°$, $p' = 16{\cdot}15$ Å. Slewing the co-ordinates gives rise to small local distortions in bond lengths and bond angles which can be corrected by a few cycles of energy minimization. The array of subunits generated by eqns (1a) to (1c) with $\tau = -36°$ is invariant to both a 5-fold rotation around the z axis and a 2-fold screw along the z axis, and can be described as having C_5S_2 symmetry (Makowski & Caspar, 1981).

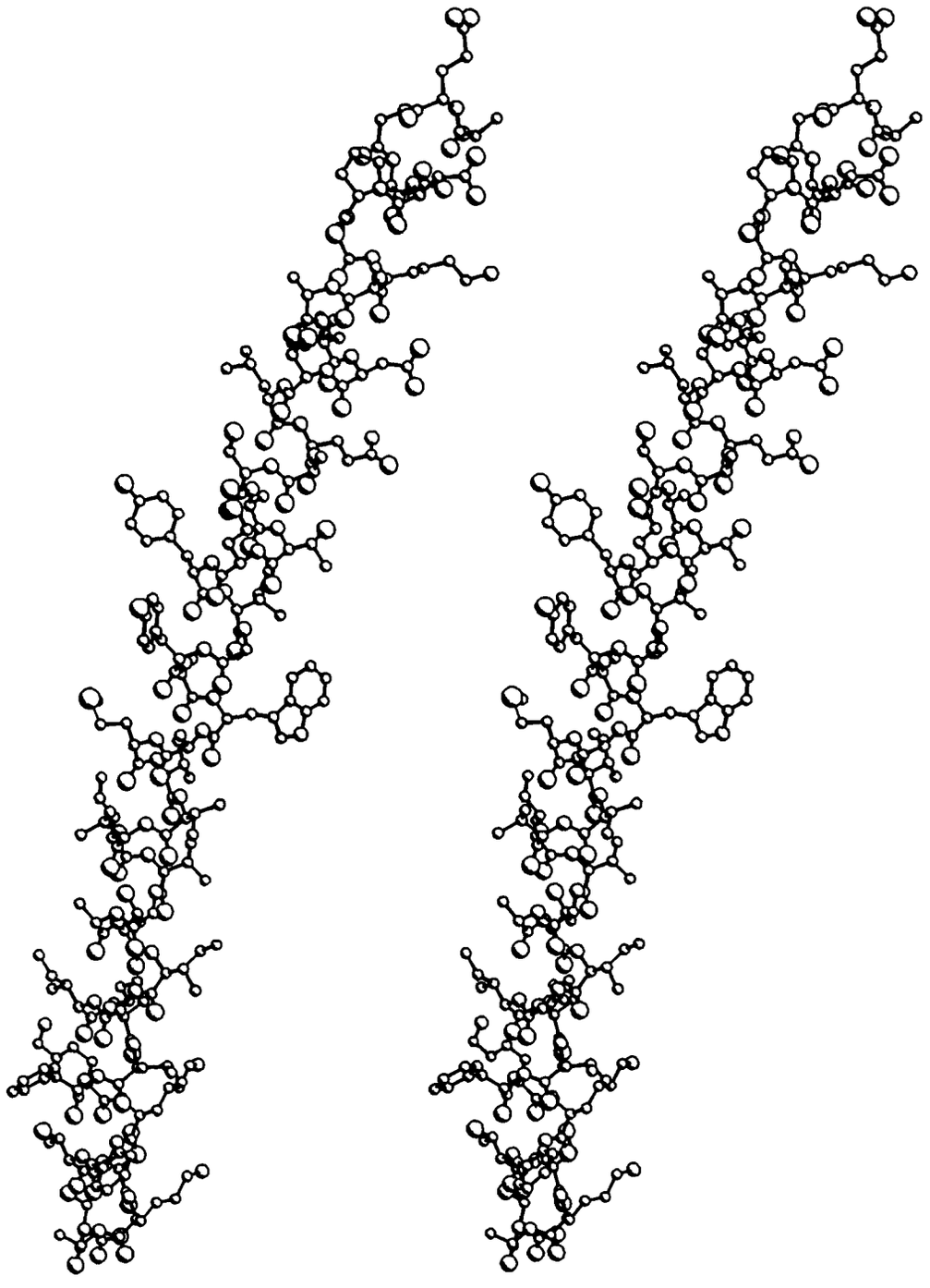
3. Results

(a) *Structural characteristics of the Ff model*

The fd^C model described by Marvin (1990) was slewed into the fd^D symmetry and refined to fit the low-resolution diffraction data, including non-diffraction constraints. The refined fd^D model was then slewed back into the fd^C symmetry and further refined by a few cycles of energy minimization to remove distortions due to slewing. The root-mean-square difference (using the method of McLachlan, 1982) between the co-ordinates of the non-hydrogen atoms in the original fd^C model 1IFD and in the refined fd^C model that we use here is 0·21 Å for the backbone and 0·35 Å for the whole subunit. The only significant differences between the two fd^C models are in the orientation of the Trp26 side-chain and in the details of the α-helix between residues 46 and 48. Each subunit of the fd^C model is represented as a single gently curved 50-residue stretch of α-helix (Fig. 1(a) and (b)). The model is consistent with the model for the protein backbone of residues 40 to 45 reported by Cross & Opella (1985) on the basis of n.m.r. experiments (Fig. 1(c)).

The interaction between nearest-neighbour subunits in the fd model is illustrated in Figure 2. Figure 2(a) represents the axes of neighbouring α-helices. The distance between axes of neighbours is similar in both the 0 to 6 and 0 to 11 directions. The crossing angle between nearest neighbours in the 0 to 11 direction is negative, but in the 0 to 6 direction the curves are nearly parallel (Table 2). The orientation of the α-helix axis changes only slightly from fd^C to fd^D, with corresponding slight changes in distances and crossing angles. Figure 2(b) shows interlocking of side-chains on neighbours. Apolar side-chains are buried and polar side-chains are exposed in the virion assembly (Table 1).

The g8p assembly forms a cylindrical hollow shell, with space in the centre for DNA, as shown in Figure 3. Figure 3(a) illustrates the indexing that we use to discuss symmetry-related subunits. Figure 3(b) and (c) show interlocking of side-chains, and the way in which basic residues are directed towards the inner surface of the protein shell. The molecular boundary used for calculation of solvent exclusion (discussed in Materials and Methods, section (b), above) was used to determine the cross-sectional area of the central hole. This area varies along the length of one 16 Å repeat unit in the virion. The mean for a set of sections through the model, calculated at intervals of 4 Å along the 16 Å repeat, is $370(\pm 35)$Å^2. Glucksman *et al.* (1992) estimated

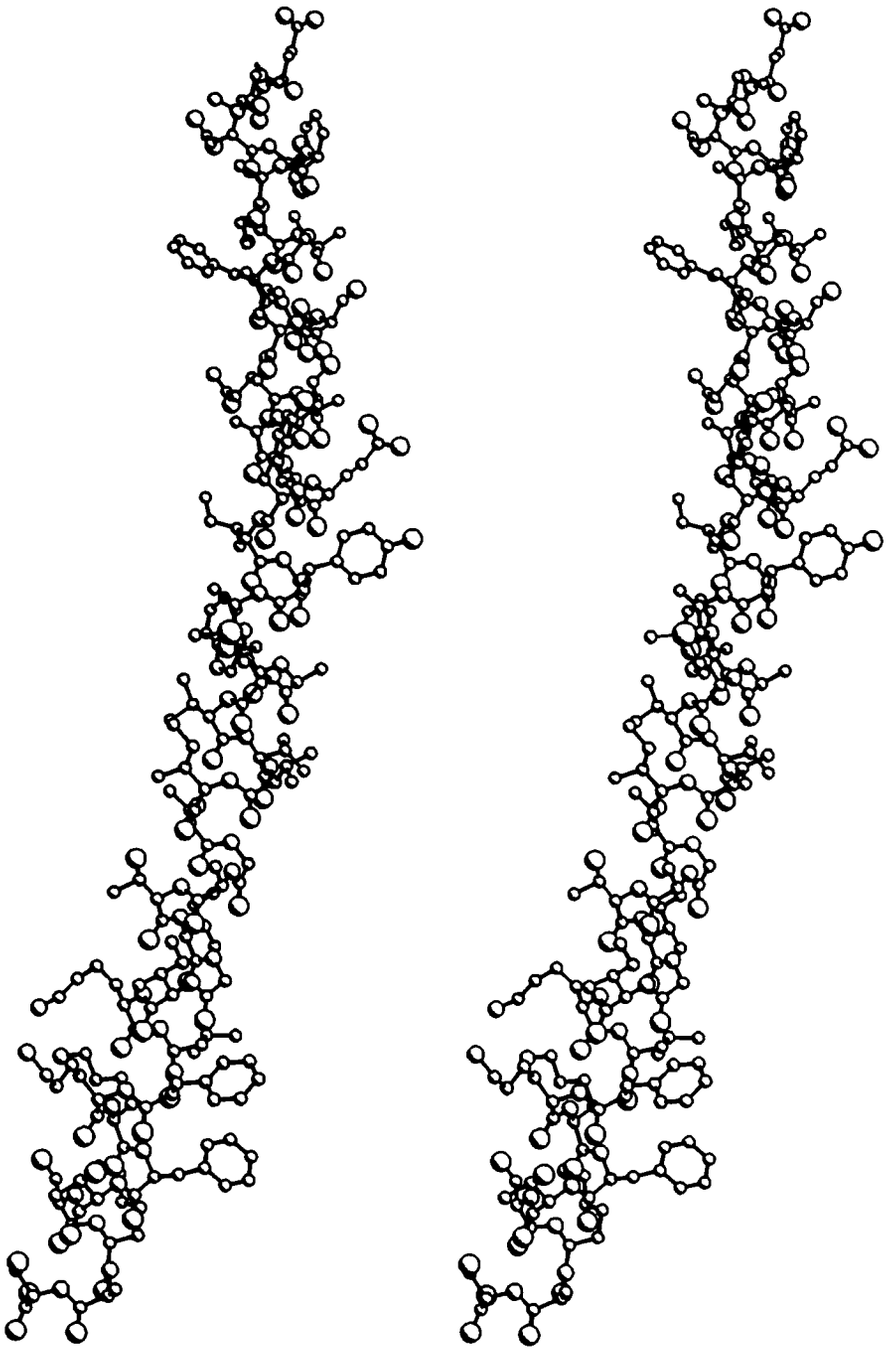


(a)

Figure 1. Molecular model of the fd^{C} α-helix subunit. Virion axis is vertical, with the N terminus of the subunit towards the top. Atom types are coded by circles whose size increases with atomic number: C < N < O < S. Atoms are shaded as if illuminated from the upper right. The stereo pairs are (a) Single subunit. View along a radius towards the virion axis, from larger to smaller radius away from the viewer. (b) View perpendicular to (a), along a tangent to a circle centred on the helix axis, from smaller to larger ϕ away from the viewer, so the outside of the virion is to the right and the inside to the left. (c) View as (a) of residues 40 to 45 (backbone atoms plus C^{β} atoms). Lighter lines show the model of Cross & Opella (1985), translated parallel to the x, y and z axes and rotated around z to superimpose their model onto our model.

from electron density maps that the area of the central hole varies along the length of one repeat from 200 to 500 Å^2.

X-ray fibre diffraction patterns of wild-type fd at neutral pH (Fig. 4(a)) show an effect (layer-line fanning) that complicates the layer-line pattern and therefore the analysis of the diffraction patterns. This effect was reported by Banner *et al.* (1981) and Bhattacharjee *et al.* (1992), and is discussed in more detail in section (b), below. There are two ways of

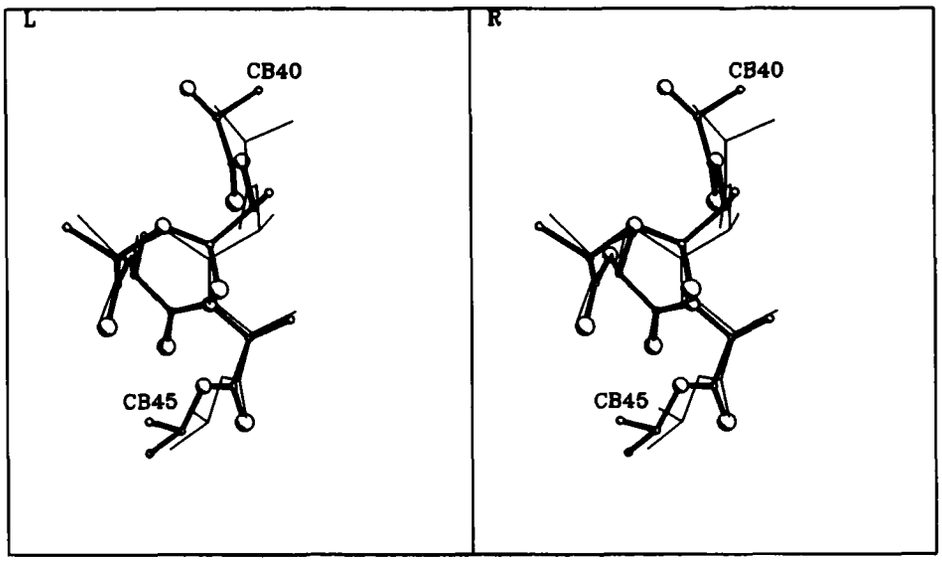


(b)

Fig. 1.

dealing with this problem to enable the analysis of the diffraction pattern, both of which involve changing the properties of the virion surface. The first way is to prepare wild-type fd fibres at a pH below the isoelectric point, which also changes the virion to the fd^D symmetry (Banner *et al.*, 1981; Bhattacharjee *et al.*, 1992). The second way is to use a mutant with an altered amino-acid residue on the outer surface of the virion, which may leave the virion near the canonical symmetry (Fig. 4(b)). Different kinds of surface exchanges have a similar effect (Bhattacharjee *et al.*, 1992), and we define the canonical Ff symmetry to be the same as that of the other class I virions (Marvin, 1990).

As illustrated in Figure 5, the Fourier transform calculated for the fd^C model is similar to the observed fd^C diffraction pattern (Fig. 4(b)) to a nominal resolution of 3·3 Å, and the calculated fd^D transform is similar to the observed fd^D pattern (Banner *et al.*, 1981; Bhattacharjee *et al.*, 1992; Glucksman *et al.*, 1992). The distribution of strong and weak intensity is the same on the calculated simulated transforms as on the observed diffraction patterns. Changing from fd^C to fd^D causes resolved



(c)

Fig. 1.

Bessel function terms to merge into one another, but does not otherwise significantly change the distribution of intensity along the layer lines (compare the 2 halves of Fig. 5).

The diffraction data used for refinement extend to a nominal resolution of 7 Å in the equatorial direction and 30 Å in the meridional direction. This region includes the strong 10 Å near-equatorial α-helix diffraction, and the fit of calculated to observed amplitudes in this region supports the lateral packing of α-helices in the model (Fig. 6(a)). The diffraction data in this region give less information about periodicities along the α-helix, but this aspect of the structure is constrained by two factors: the side-chain interlocking in the model and the fit of calculated to observed data on higher layer lines (Fig. 5).

The calculated transform on $l = 1$ of the rod

Table 2

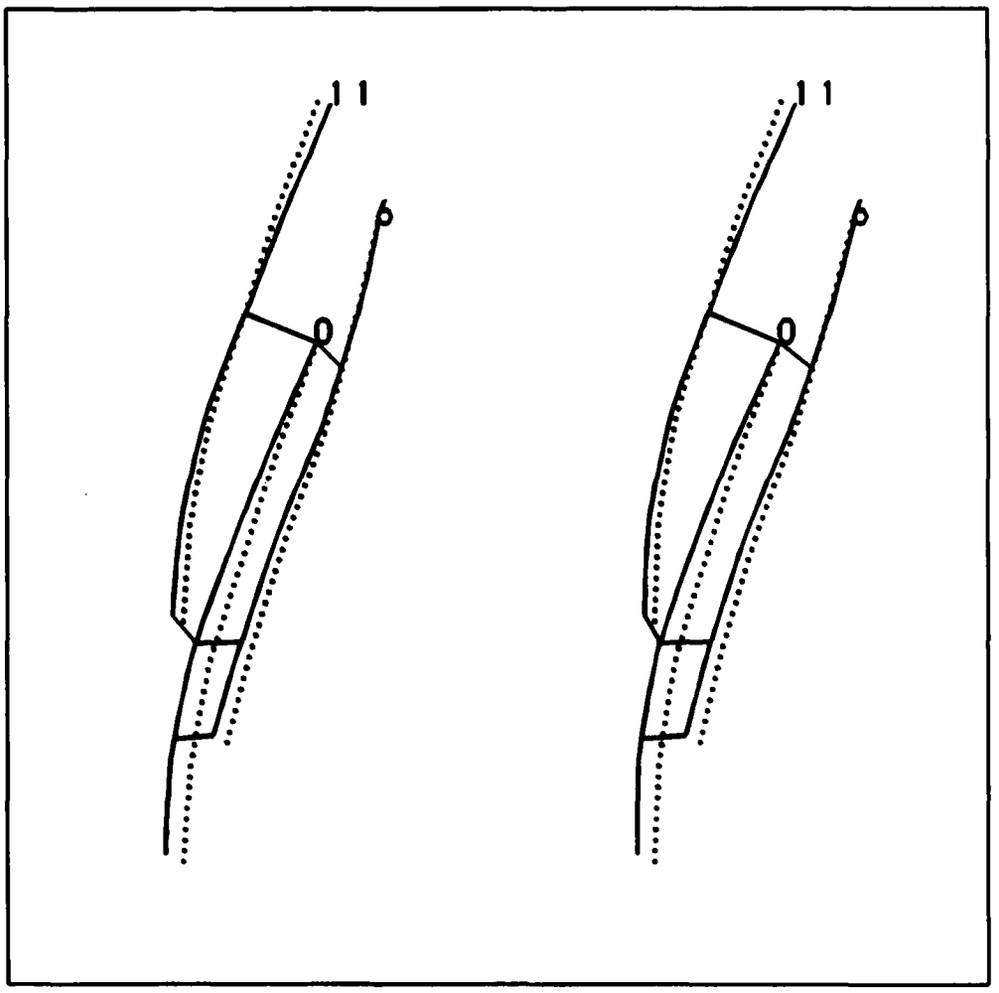
Comparison of the fd^C and fd^D models

	fd^C		fd^D	
A. *Properties of the α-helix axis*†				
κ(rad/Å)	0·0065(±0·0010)		0·0044(±0·0015)	
τ(rad/Å)	0·0146(±0·0029)		0·0112(±0·0041)	
B. *Contacts of the axis to neighbours*‡				
	Distance (Å)	Angle (°)	Distance (Å)	Angle (°)
		$k = 0$ to $k = 6$		
Top	10·4	−5	10·7	−5
Middle	9·5	−4	9·7	−2
Bottom	9·1	5	9·0	3
		$k = 0$ to $k = 11$		
Top	10·4	−18	10·6	−16
Bottom	10·7	−17	10·2	−15
C. *Properties of the protein subunits*§				
ϕ_i(deg.) ($i = 2 - 49$)	−63·4(±3·8)		−65·5(±5·8)	
ψ_i(deg.) ($i = 2 - 49$)	−40·3(±5·1)		−38·6(±6·1)	
ω_i(deg.) ($i = 1 - 49$)	179·8(±2·1)		180·8(±3·2)	
O_iN_{i+4}(Å)	3·09(±0·27)		3·06(±0·28)	
$C_iO_iN_{i+4}$(deg.)	152(±6)		152(±6)	
χ^1 (deg.)	−73(±16)		−73(±15)	

†Curvature (κ) and torsion (τ) were calculated for the segmented approximation to the helix as described by Marvin (1990). The units of κ and τ used by Marvin (1990) should be rad/Å, as here, and not turn/Å, as given there.

‡Contacts from the axis of subunit $k = 0$ to $k = 6$ or $k = 11$ are measured at the straight line segments shown in Fig. 2(a). Crossing angle is defined using the convention of Chothia *et al.* (1981), that "the angle is negative if the near helix is rotated in a clockwise direction relative to the far helix".

§Subscript i is position number in the sequence (Table 1). The residues used to calculate the mean χ^1 value are discussed in the text.

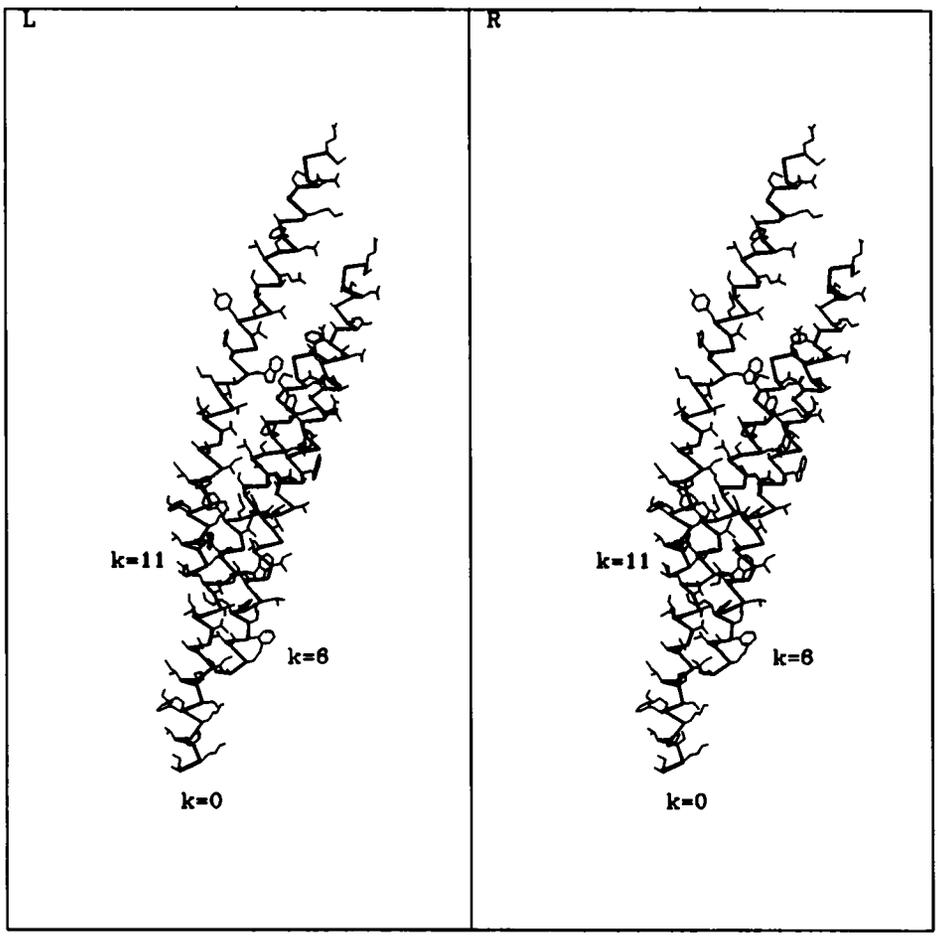


(a)

Figure 2. Nearest-neighbour interactions between subunits. View along a radius from outside the virion, towards the virion axis, as Fig. 1(a). Three subunits are shown. The central (lowest) subunit has index $k = 0$; to the right is the neighbour with $k = 6$; to the left is the neighbour with $k = 11$. (a) Axes of the subunits, generated by fitting a segmented helix approximation (Marvin, 1990) to the α-helix, with 5 segments of 10 residues each. Continuous curves, fd^C; broken curves, fd^D. The fd^D model has been rotated around z and translated parallel to z so that the (ϕ, z) co-ordinates at the N terminus of $k = 0$ coincide with those of the fd^C model. The local distances between the axes and the torsion angles between axes are measured at the points on the fd^C curves joined by straight line segments, and at corresponding points on the fd^D curves, and are listed in Table 2. (b) Interactions between α-helix subunits. Heavy lines join C^α atoms of the fd^C model, and lighter lines join the atoms of the side-chains.

model of Glucksman *et al.* (1992) is remarkably similar to that of the α-helix model (without side-chains) of Banner *et al.* (1981), even though the two models have subunit axes of opposite hands. At this resolution, neither the difference between a helix and a rod nor the enantiomorph can be distinguished using native data alone. The transforms of both these models fit the observed data much less well than the transform of our model with side-chains (Fig. 6(a)), even though our model is constrained to fit chemical data as well as X-ray data. The poor fit of models without side-chains can be traced to the minimum in the radial electron density distribution caused by the girdle of apolar side-chains in the interior of the *Inovirus* protein shell (Marvin & Wachtel, 1976), which is not taken into account if side-chains are omitted.

In unpublished experiments (cited by Marvin, 1990), Dr R. S. Brown applied to fd the methods for iodination that he developed for Pf1 (Nave *et al.*, 1981) and found that both Tyr21 and Tyr24 are accessible to iodination. The fact that Tyr21 is not on the surface of the model (Fig. 1(b)) might seem inconsistent with this observation, but the flexible N-terminal region proposed by Opella *et al.* (1987) and discussed in section (c) (iii), below, as an explanation for the partial accessibility of Trp26 could also make Tyr21 accessible to iodination. To attach iodine atoms to the tyrosine residues in the model, we replaced the bonds linking $C^{\varepsilon 1}$ and $C^{\varepsilon 2}$ to hydrogen by bonds of length 2·05 Å to I. Trial models show that only small adjustments of the fd model are necessary to accommodate the two iodine atoms on Tyr21, and Tyr24 on the surface of the



(b)

Fig. 2.

virion can be iodinated without difficulty.

The amplitude differences observed by X-ray fibre diffraction of M13 with and without iodination (Glucksman *et al.*, 1992) are similar to those predicted by our model (Fig. 6(b)) and further support it. We assume with Glucksman *et al.* (1992) that both Tyr are di-iodinated. The calculated amplitude at about 0·04 $Å^{-1}$ on $l = 1$ is greatly increased by attaching iodine to Tyr21 and Tyr24 on the model (Fig. 6(b)) as observed experimentally (Glucksman *et al.*, 1992), and there are further maxima at 0·063 $Å^{-1}$, 0·088 $Å^{-1}$ and 0·107 $Å^{-1}$ on this layer line of the calculated iodinated transform, roughly corresponding with the observed amplitude changes. The differences are less dramatic on $l = 2$ and $l = 3$, but the qualitative consequence of iodination is similar for the model and for the observed data. For instance on $l = 2$, the iodinated amplitude is higher than the native around 0·03 $Å^{-1}$ and 0·08 $Å^{-1}$ for both the model calculations and the observed data; and on $l = 3$ the iodinated amplitude falls below the native at about 0·04 $Å^{-1}$ and is above the native in the region 0·10 to 0·12 $Å^{-1}$ for both calculated and observed data.

The native data for M13 on $l = 1$ presented by Glucksman *et al.* (1992) differ somewhat from those of fd^D measured by Banner *et al.* (1981) that we present in Figure 6(a). These differences probably reflect not the differences between the structures of the two virions, but instead the difficulty of extracting an accurate molecular transform from experimental fibre diffraction data. There is fine structure in the data of Glucksman *et al.* (1992), in particular a significant minimum in amplitude at about 0·10 $Å^{-1}$ on $l = 1$ not observed by Banner *et al.* (1981). The low-order Bessel function terms predicted on $l = 1$ for a single Ff virion of 35 Å maximum radius cannot have the rapid fluctuations in amplitude that are implied by this fine structure. There may be uncorrected interference effects or crystal sampling in the data.

Some properties of the fd^C and fd^D models are compared in Table 2. The mean O_i to N_{i+4} hydrogen-bond length and the mean $C_iO_iN_{i+4}$ angle are within the range found in surveys of α-helices (Baker & Hubbard, 1984). Backbone torsion angles (ϕ, ψ) are within the standard α-helix range ($-65(\pm 12)°$, $-39(\pm 11)°$) cited by Morris *et al.* (1992), and are essentially unchanged between the two structures. The slight change in curvature and torsion of the α-helix axis does not introduce any significant distortion into the α-helix. The only

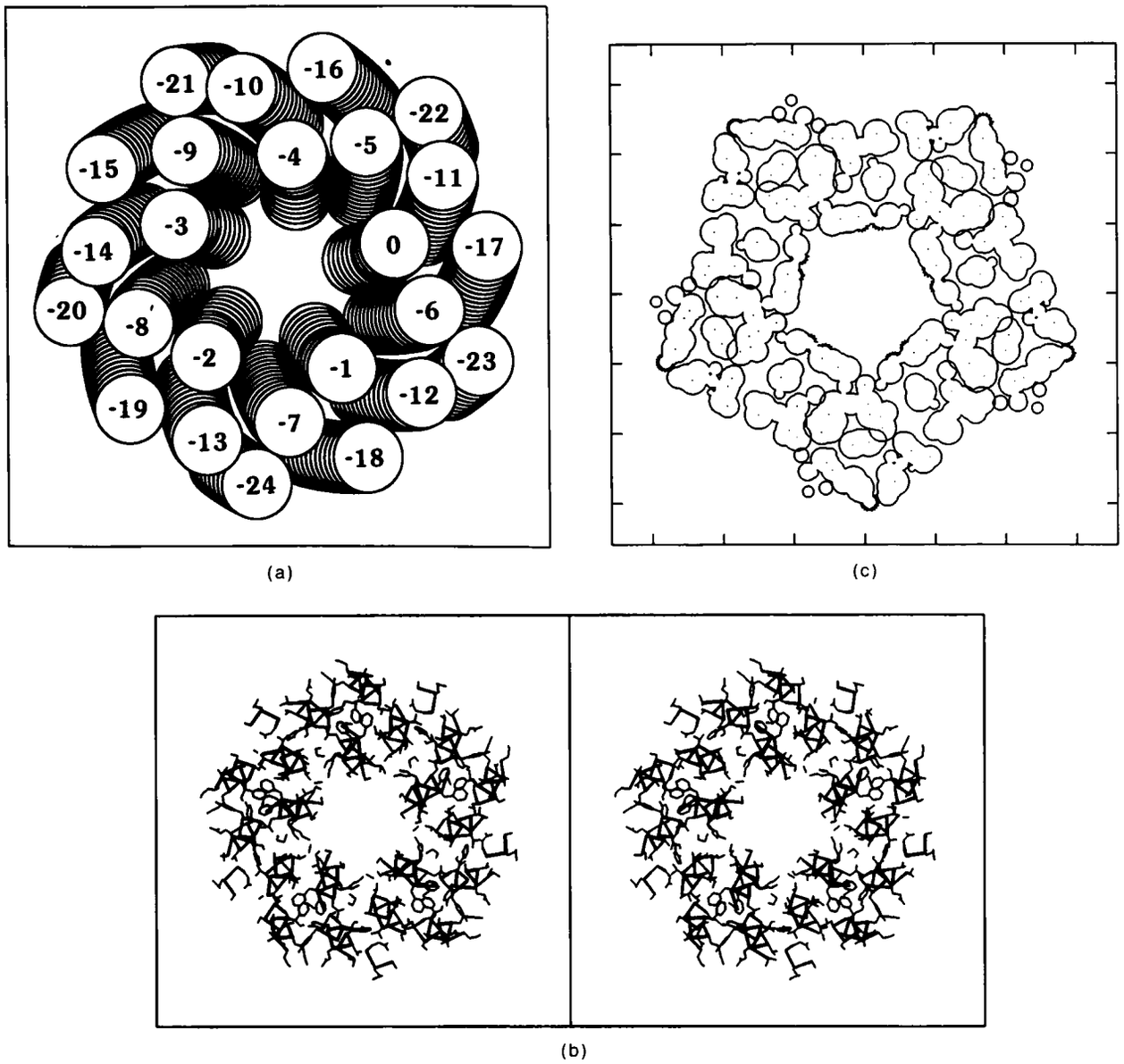
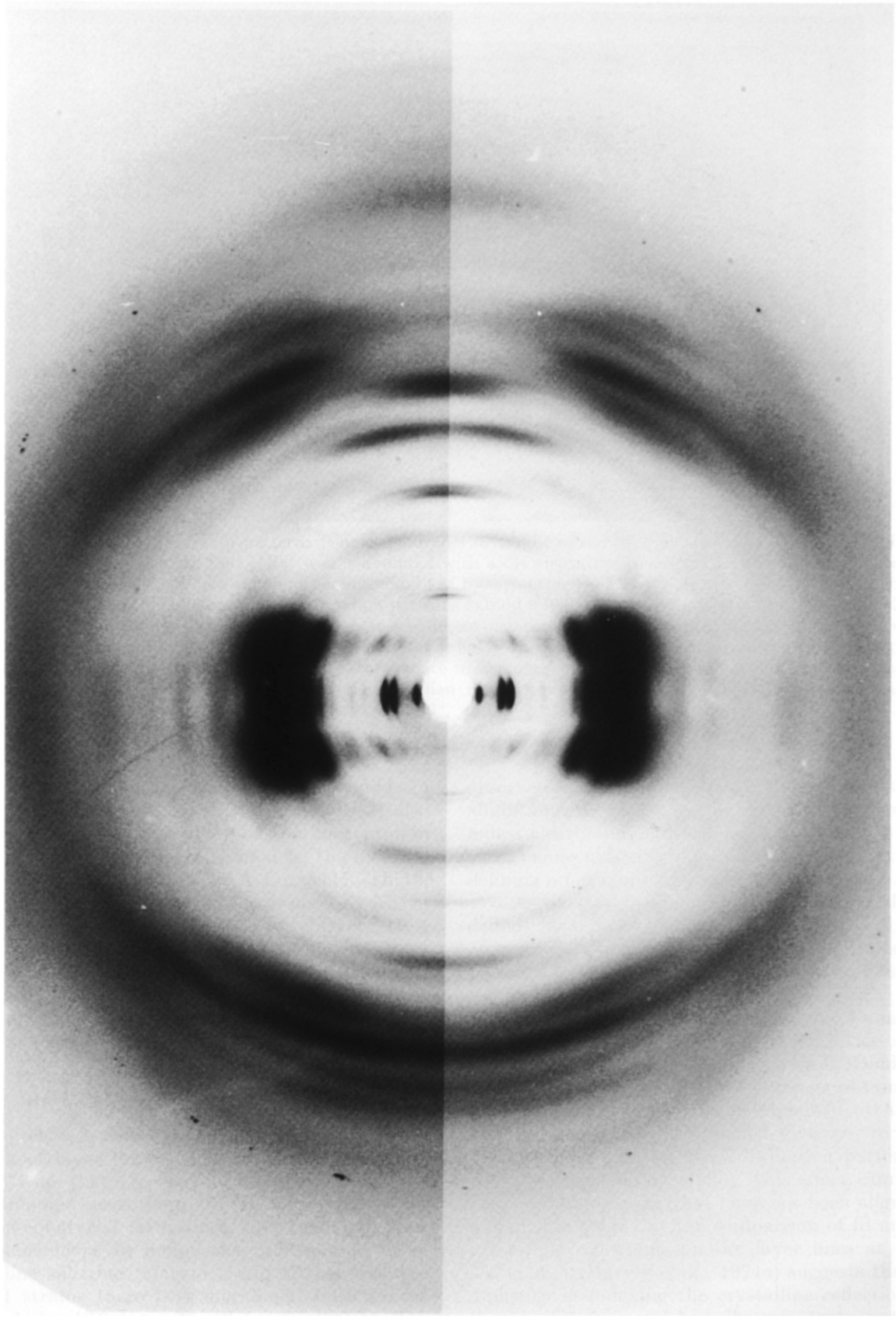


Figure 3. Model of the protein coat of the fd virion. View direction is parallel to the virion axis, from larger towards smaller z away from the viewer (i.e. from smaller towards larger residue number in the amino-acid sequence). A full array of symmetry-related copies of the subunit was generated by the operation of the virion helix parameters. (a) A slab about 20 Å thick was cut from the array of subunits, with subunits represented schematically by a stack of disks 9 Å in diameter and 1·5 Å thick. The indices k of the symmetry-related subunits are shown within each subunit, as calculated from eqn (1). Subunits with origins at smaller z than the reference ($k = 0$) subunit, corresponding to $k < -4$, are cut by this slab at larger radius than the reference subunit (compare Marvin *et al.*, 1974*a*, Fig. 10). (b) Slab as (a), with the backbone of the subunit represented by heavy lines connecting C^{α} atoms, and with lighter lines connecting the side-chains atoms. (c) Section perpendicular to the virion axis, through the van der Waals outline of the fd protein coat. Fiducial marks are at intervals of 10 Å. Centres of all atoms within $\pm 0{\cdot}5$ Å of the section are shown as points. Group radii (including hydrogen atoms) are used instead of atomic radii (Richards, 1985) so overlaps of non-bonded circles should not be taken to indicate poor van der Waals contacts. The basic residues are shown with jagged van der Waals outlines. The section is taken at a z value between the N^{ζ} of residues Lys43 and Lys44. The 5 symmetry-related copies of these 2 atoms are seen (jagged) on the inside of the shell, and N^{ζ} of Lys8 is seen on the outside of the shell.

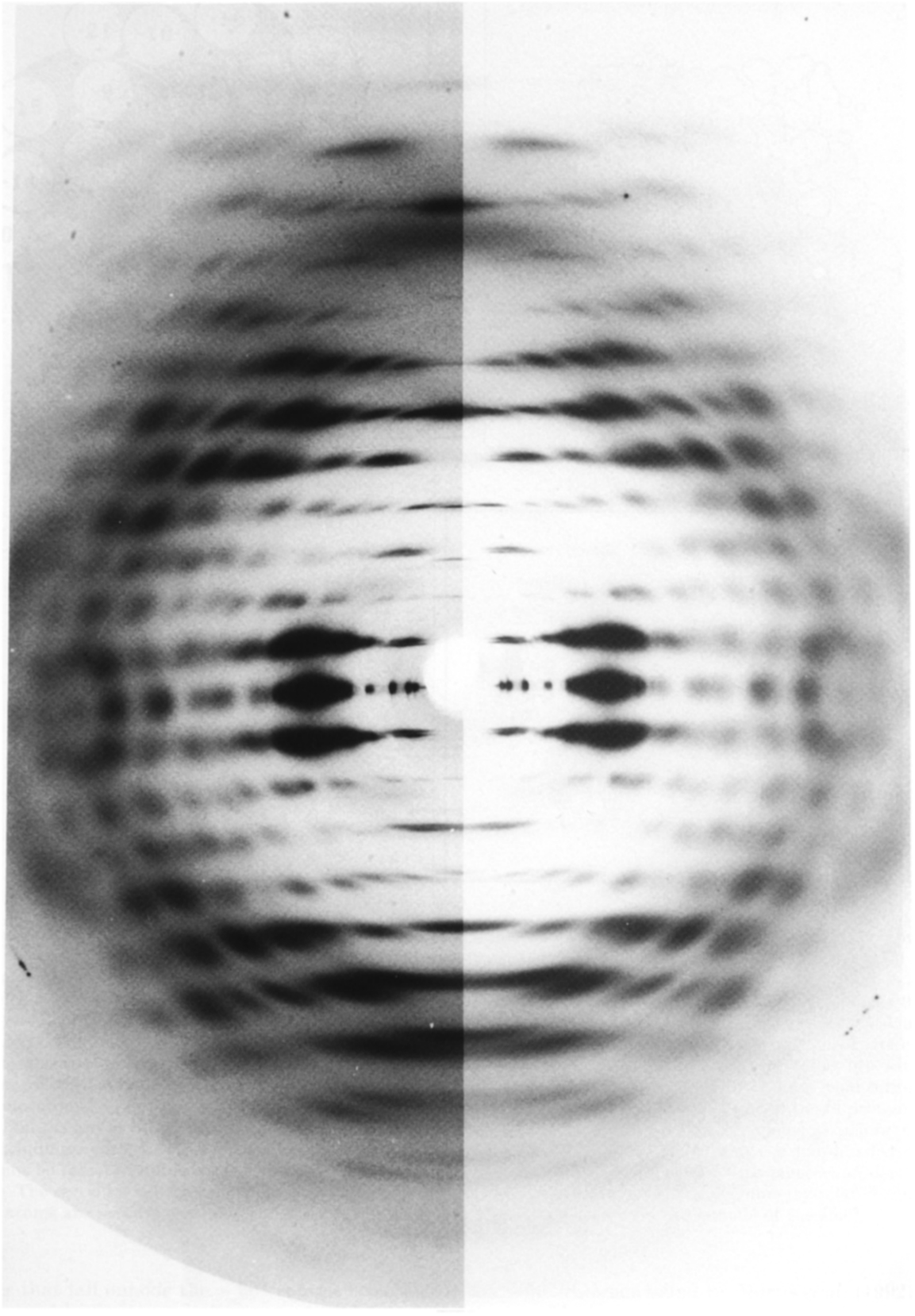
residues that fall outside the $-60°$ rotamer rule for side-chain χ^1 torsion angles are Asp4 ($+60°$ rotamer) and Gln15 and Met28 (180° rotamer), which are also acceptable rotamers for α-helix. The mean value of the remaining 32 χ^1 angles differs little between fd^C and fd^D (Table 2), and is similar to the mean value $\chi^1 = -67(\pm 15)°$ for side-chains in the $-60°$ rotamer found by Morris *et al.* (1992) in a survey of high-resolution protein crystal structures.

Ser and Thr residues at position i in an α-helix often form hydrogen bonds to the backbone carbonyl oxygen atom on residue $i - 3$ or $i - 4$ on the same helix when χ^1 is in the $-60°$ rotamer (Baker & Hubbard, 1984; Gray & Matthews, 1984).



(a)

Figure 4. X-ray fibre diffraction patterns of Ff. The meridian (fibre axis direction) is vertical, but the fibre has been tilted by about 13° away from perpendicular to the X-ray beam to bring the diffuse 3·4 Å meridional intensity (top) onto the sphere of reflection. The right half of each photograph has been printed lighter to reveal more detail in the regions of very strong diffracted intensity. (a) Diffraction pattern of wild-type fd. (b) Diffraction pattern of f1g8Y21M.



(b)

Fig. 4

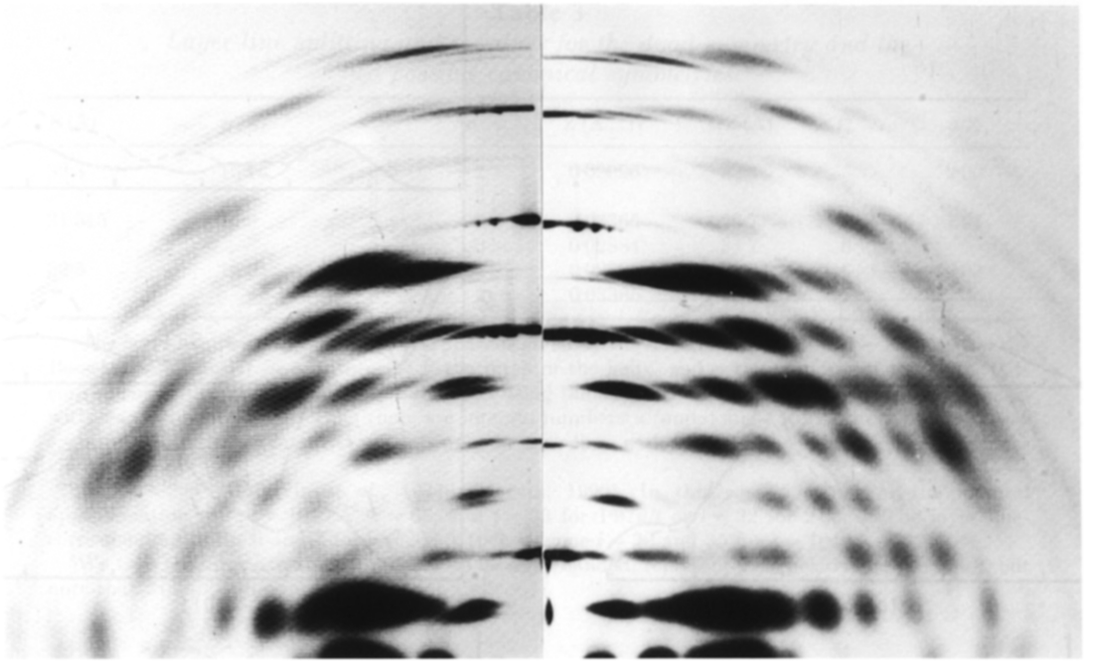


Figure 5. Simulated diffraction patterns (single quadrant) of the fd model. The calculated water-weighted Fourier transform of the protein coat $\Sigma_n|G_{nl}(R)|^2$ was convoluted with a disorientation function corresponding to 3° standard deviation of disorientation, and a coherent particle length of 1000 Å, using the CIN subroutine (Provencher & Glöckner, 1982; Marvin *et al.*, 1987). The simulated diffraction patterns extend to 0·3 $Å^{-1}$ in the equatorial direction and 0·35 $Å^{-1}$ in the meridional direction. Bessel function orders to $n = 30$ are included; increasing the order gave no detectable change in the diffraction pattern. The simulated patterns are at roughly the same magnification as the experimental pattern (Fig. 4), but have not been converted from reciprocal space to film space, and are therefore not precisely comparable to the experimental pattern. Sharp spots near the meridian result from flaws in the display program, and are not significant. Left, canonical symmetry, fd^C: $\tau = -33{\cdot}23°$, $p = 16{\cdot}00$ Å. Right, diad symmetry, fd^D: $\tau' = -36{\cdot}00°$, $p' = 16{\cdot}15$ Å.

The argument favouring these hydrogen bonds in membrane proteins (Gray & Matthews, 1984) also applies in the apolar interior of the coat protein shell. In both the fd^C and fd^D models, the distance from O_i^γ to O_{i-4} is 2·8 to 3·0 Å for Ser17, Thr19, Thr36 and Ser50, consistent with a good hydrogen bond. Ser13, Thr46 and Ser47 have O_i^γ to O_{i-4} distances of 3·0 to 3·4 Å, but these residues are exposed to solvent (Table 1) and competition with hydrogen bonds to water is likely (Barlow & Thornton, 1988). All O_i^γ to O_{i-3} distances for the Ser and Thr residues are greater than 3·5 Å.

(b) *Polymorphism of class I virions*

Some class I diffraction patterns have a typical pattern of layer lines which can be considered as "split" from that expected for a perfect screw dyad axis. In some cases there are crystalline reflections on non-equatorial layer lines, and indexing these reflections helps to define the precise amount of layer line splitting (Marvin *et al.*, 1974*a*). For the class I strains there is a meridional reflection at about $1/Z = 16{\cdot}0$ Å (or $l = 13$ for $c = 208$ Å). The value of this meridional spacing depends on the hydration of the fibre, and for a 16·0 Å meridional spacing, the hexagonal unit cell dimension defined by the equatorial reflections is about $a = 57$ Å (Dunker *et al.*, 1974). For some fibres with this degree of hydration, crystalline reflections are observed on the non-equatorial layer lines with $1/Z$ about 32 Å. For If1, these non-equatorial reflections index clearly on a larger hexagonal lattice with three molecules passing through the unit cell, defining a hexagonal unit cell with dimension about $a' = 99$ Å (Marvin *et al.*, 1974*a*). This indexing defines layer lines split slightly from 1/32 Å, with $1/Z = 34{\cdot}7$ Å and 29·7 Å ($l = 6$ and 7). The strong 10 Å intensity is observed on the layer line with $1/Z = 29{\cdot}7$ Å ($l = 7$).

For fd the indexing is less clear, but we use the same canonical symmetry for Ff and IKe as for If1. Unpublished measurements with C. Dohle and E. Marseglia suggest that fd patterns obtained at neutral pH (Fig. 4(a)) can be explained by a skew pseudo-hexagonal unit cell which gives rise to layer-line "fanning", as for fibre diffraction patterns of collagen (Fraser *et al.*, 1983). This effect cannot be seen clearly for fibres that have not been aligned in a magnetic field. But the comparison of fd and If1 crystalline reflections on the layer lines at about $1/Z = 32$ Å (Marvin *et al.*, 1974*a*) suggests that the difficulty in indexing the crystalline reflections on the fd patterns may have been a consequence of layer line fanning: that is, fanning is a property of fd, not of the magnetic field. This interpretation is supported by consideration of the liquid crystal properties of fd (Booy & Fowler, 1985). Bhattacharjee *et al.* (1992) proposed helical aggregates to explain these diffraction patterns.

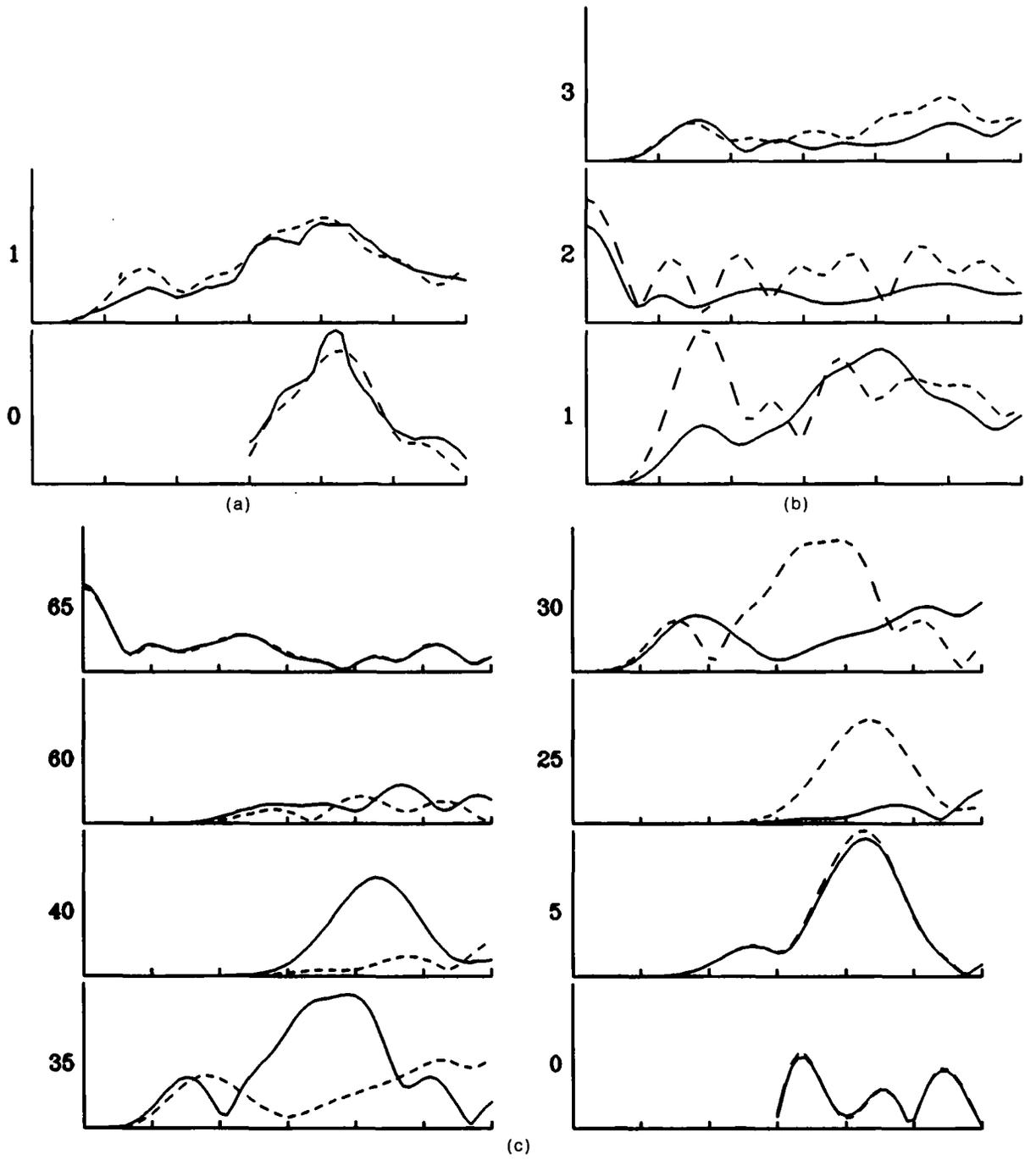


Figure 6. Low resolution Fourier-Bessel transforms of fd models showing the strong 10 Å near-equatorial region. Curves correspond to 1 quadrant of the fibre pattern, with the meridian vertical and the equator horizontal. Each curve represents 1 layer line. The index l is shown to the left of the layer line. Bessel function orders to $n = 15$ are included; increasing the order did not add appreciable calculated intensity. Amplitude $|G_n|$, not intensity, is plotted. Where more than one Bessel function order contributes to the layer line, the amplitude $|G_n|$ calculated for each Bessel function order was squared, the contributions were summed, and the square root of the sum was plotted. Horizontal scale divisions along $R = 2\sin\theta/\lambda$ are at intervals of 0·025 $Å^{-1}$ from $R = 0$ at the left-hand side of each curve. Data for $R < 0{\cdot}075$ $Å^{-1}$ are not included on $l = 0$ because the DNA contributes substantially to the observed diffraction in this region (Wachtel *et al.*, 1974), and our model of fd DNA has not been refined. The DNA is about 12% by weight of the virion, but it does not have the same helix symmetry as the protein coat and does not contribute to the protein layer lines except on the equator. Therefore we omit the DNA from our calculations. (a) Model of fd^D ($2{\cdot}00\,u/t$). The 4 layer lines for $l = 25$, 30, 35 and 40 in (c) coalesce for this symmetry onto a single layer line $l = 1$ for $c = 32{\cdot}3$ Å. Observed amplitudes $|I_l(R)|^{\frac{1}{2}}$ from Banner *et al.* (1981) are shown as continuous curves. Calculated amplitudes $|G_{nl}(R)|$ are shown as broken curves. The correlation coefficient between observed and calculated amplitudes on $l = 1$ is 0·96. (b) Model as (a). Continuous curves, native calculated transform; broken curves, transform of model with di-iodinated Tyr21 and Tyr24. (c) Continuous curves, fd^C ($1{\cdot}97u/t$, Table 3), showing that the strong 10 Å amplitude (at about 0·1 $Å^{-1}$) appears on $l = 35$ for $c = 1040$ Å. Broken curves, model of fd with $2{\cdot}03u/t$, showing that for this option, strong 10 Å intensity is predicted on $l = 30$, not $l = 35$.

Table 3
Layer-line splitting and n *values for the dyad symmetry and the two possible canonical symmetries*

P (Å)	p (Å)	n	m	Z (Å^{-1})†	$1/Z$ (Å)	l‡	u/t§
32·3	16·15	5	−2	0·03096	32·3		2·00
—	—	−5	3				
31·515	16·0	5	−2	0·03365	29·7	7	1·97
—	—	−5	3	0·02884	34·7	6	
32·5	16·0	5	−2	0·02884	34·7	6	2·03
—	—	−5	3	0·03365	29·7	7	

†The layer line position Z in reciprocal space is given by the selection rule $Z = n/P + m/p$, with n the Bessel function order, m any integer, P the pitch of the helix, and p the unit rise of the helix. It is convenient to define a rational approximation using the smallest integers u and t such that $p/P = t/u$ within experimental error. Then there are integral numbers u (units) and t (turns) in the helix repeat $c = up = tP$, diffracted intensity appears only on layer lines with $l = Zc$, and the selection rule becomes $l = tn + um$.

‡For $c = 208$ Å (Marvin *et al.*, 1974*a*; Marvin, 1990). In this case $u = 13$. For the integral approximation of t, $c = 1040$ Å, $u = 65$; and $t = 33$ for $1{\cdot}97u/t$ or $t = 32$ for $2{\cdot}03u/t$. The layer lines corresponding to $l = 6$ and 7 for $c = 208$ Å then become $l = 30$ and 35 for $c = 1040$ Å.

§We take the approximate 2-fold screw to be right-handed, represented as positive values of u/t, but note that $2{\cdot}03u/t$ is identical to $-1{\cdot}97u/t$.

At a pH below the isoelectric point, the fd layer line fanning disappears, and the helix symmetry changes slightly so the two layer lines merge into one at $1/Z = 32{\cdot}3$ Å (Banner *et al.*, 1981) with a meridional reflection at 16·15 Å. The five subunits per 16·15 Å repeat are related by an exact 2-fold screw axis, with $u/t = 2{\cdot}00$. The same slight twist to an exact screw dyad symmetry can be induced by genetic modification of the charge on the surface of the virion at neutral pH, for instance by a D12N exchange (strain M13, Bhattacharjee *et al.*, 1992). The canonical pattern at neutral pH can then be understood as a slight twist away from a perfect dyad, giving rise to layer line splitting (Table 3). But which way is the twist: to u/t slightly less or slightly more than 2·00? To determine this we need to know the orientation of the α-helix within the virion. In our models, the α-helix axis is right-handed; this choice is supported by calculations for heavy-atom derivatives (Glucksman *et al.*, 1992 and Fig. 6(b)). There will be strong intensity for Bessel functions of order n on a layer line at position Z only if the argument $z_jZ - n\phi_j$ of the exponential is constant for most of the j atoms in the subunit (see eqn (6) of Marvin *et al.*, 1987). That is, $n/Z = z_j/\phi_j$. Since the layer line spacing Z is positive, and for a right-handed helix z_j/ϕ_j is positive, n must be positive. The strong intensity in the 10 Å region is indexed on $1/Z = 29{\cdot}7$ Å, not $1/Z = 34{\cdot}7$ Å (Fig. 4(b); see also Marvin *et al.*, 1974*a*), so $u/t = 1{\cdot}97$, not 2·03 (Table 3). This was the reasoning that led to the choice of the canonical symmetry by Marvin (1990). This type of argument is discussed in more detail by Marvin & Nave (1982).

The question of whether the canonical symmetry is slightly less or slightly more than 2·00 u/t can be illustrated by calculating transforms of molecular models for the two options. These transforms are shown in Figure 6(c) for fd^C with 1·97 u/t and for the (rejected) 2·03 u/t symmetry.

(c) *Comparison of class I wild-type and mutant strains*

(i) *Side-chain environment in class I models*

The environment of the amino-acid side-chains in a protein model is a useful measure of the validity of that model. In the method of three-dimensional profile analysis (Bowie *et al.*, 1991; Lüthy *et al.*, 1992), the environment of each side-chain is defined by three parameters: the surface area of the side-chain that is masked from solvent by other parts of the structure, the fraction of the side-chain surface area that is covered by solvent or by polar residues on other parts of the structure, and the local secondary structure. The probability of finding a given type of side-chain in a given environment has been defined (Bowie *et al.*, 1991) by analysis of a large database of known three-dimensional structures, so any new model can be tested against these probabilities. The environments of side-chains in our canonical class I models are given in Table 1. Most side-chains in the apolar domain (positions 21 to 39) are buried in the virion assembly, and apolar side-chains nearer the two ends of the subunit are also generally buried. Pro6, Ala9 and Ala16 are all relatively accessible in fd, but these positions are occupied by polar and/or charged residues in If1 and/or IKe (Table 1). The profile scores of the canonical models are 26 for fd, 29 for If1 and 23 for IKe. These values are all in the range expected for proteins of about 50 residues, as extrapolated from highly refined structures (Lüthy *et al.*, 1992).

Profile window plots of the models are illustrated in Figure 7. The profile score of the fd protein in the virion assembly is substantially higher than the score of the isolated subunit along most of its length (Fig. 7(a)). This supports the calculation of Marvin (1990), that the protein subunit has a lower solvation free energy in the virion than in the monomer. The profile score of the virion assembly along the

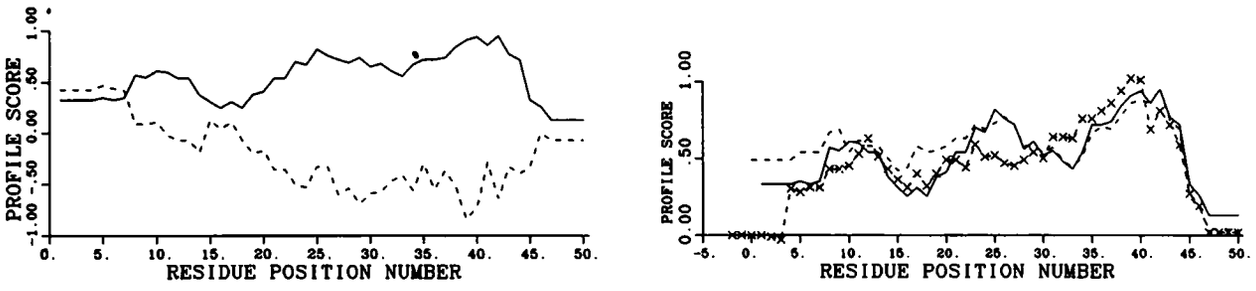


Figure 7. Profile window plots for class I models. The vertical axis gives the average score for residues in a 7-residue sliding window, the centre of which is at the residue position (defined in Table 1) indicated by the horizontal axis. Scores for the first 3 and the final 3 positions have no meaning. (a) Profile for the fd protein subunit surrounded by its nearest neighbours in the fd^D symmetry (continuous curve) and for the isolated fd subunit (broken curve). (b) Profile for the class I protein subunit surrounded by its nearest neighbours in the canonical symmetry. Continuous curve, fd; broken curve, If1; broken curve with symbol (X), IKe.

length of the sequence is generally as high as for the correct model of various small proteins, and higher than for incorrect or misfolded models (Lüthy *et al.*, 1992). There are exceptions at the two ends. At the N terminus the profile tends to be low; the score of the first few residues at the N terminus of the fd subunit in the assembly is actually *lower* than the score for the single isolated subunit (Fig. 7(a)). This effect would be consistent with a tendency for the N-terminal region to bend away from the body of the virion (Opella *et al.*, 1987). The profile scores of the class I models are also low at the C terminus, especially for the basic residues at positions 44 and 48, but this can be attributed to the absence of the DNA core in our models. The fd^C and fd^D profiles are essentially indistinguishable (compare Fig. 7(a) and (b)).

(ii) *The inner surface of the protein shell*

The array of α-helix subunits forms a cylindrical hollow shell with space in the centre for DNA (Fig. 3). The inner surface of the protein shell is remarkable for the concentration of Lys residues facing towards the DNA (Fig. 8), where they can function to neutralize the DNA negative charge. They probably also interact with the C terminus of the α-helix dipole (Scholtz & Baldwin, 1992). Replacements allowed in the wild-type class I strains If1 and IKe are shown in Table 1. Lys48 in fd can be replaced by a neutral residue using genetic engineering, and these charge-change mutants are about a third longer than wild-type, whereas if Lys48 is replaced by Arg (thus conserving the charge), the length of the virion is unchanged (Hunter *et al.*, 1987). This observation is consistent with the proposal that the positive charges on the C-terminal end of the protein neutralize non-specifically the negatively charged phosphodiester bonds of the DNA molecule by matching the linear charge density of the negatively charged DNA rod and the positively charged protein sheath (Marvin & Wachtel, 1976).

The S47K exchange is lethal, probably because the DNA already fills the core of the virion and cannot be further compressed to match the increased positive charge density on the protein. Hybrid virions containing a mixture of S47K and K48Q subunits can be produced, however, and their length (i.e. their DNA/protein ratio) depends on the ratio of the two types of subunit (Greenwood *et al.*, 1991). Figure 8 shows that Ser47 can be replaced by

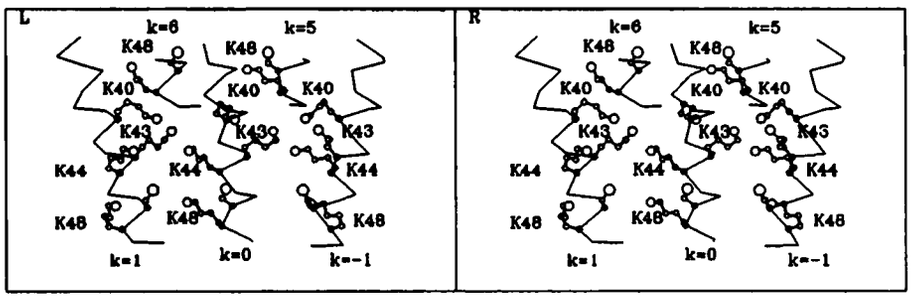


Figure 8. Basic residues near the C terminus of the fd^C model. Virion axis vertical. View from inside the virion along a radius. The α-helix backbone is shown as a thin line connecting successive C^α atoms, and side-chains are shown as heavy lines. Atom types are coded by size as in Fig. 1. The reference ($k = 0$) subunit is in the centre, showing residues K40, K43, K44 and K48 (labelled) and S47 (not labelled); the other residues in this segment of α-helix have been removed for clarity. To the left and right of the $k = 0$ subunit at the same level are the $k = 1$ and -1 subunits, showing the same sets of residues. Also to the left and right of the $k = 0$ subunit are the $k = 6$ and 5 subunits; these subunits are 16 Å higher than the $k = 0$ subunit, so only residues S47 and K48 appear in this view.

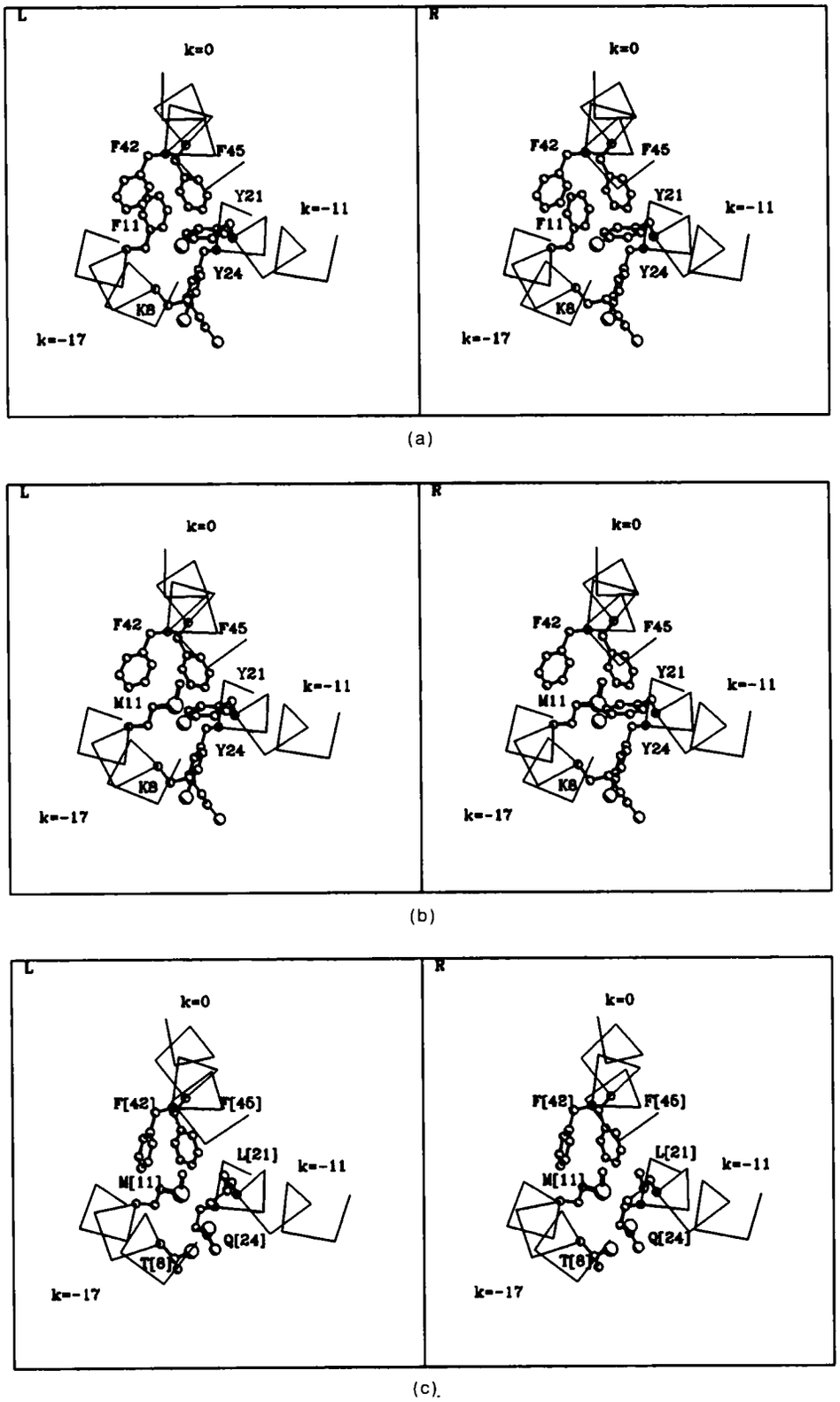


Figure 9. Intercalation of Phe residues in the model. View parallel to the axis of the virion, showing parts of units $k = 0$, -11 and -17. All side-chains except those labelled have been removed for clarity. The α-helix backbone is shown as a thin line correcting successive C^{α} atoms, and side-chains are shown as heavy lines, with atom types coded by size and shaded as in Fig. 1. (a) Wild-type fd^{C}. (b) Model of hypothetical fdg8F11M. (c) Wild-type IKe.

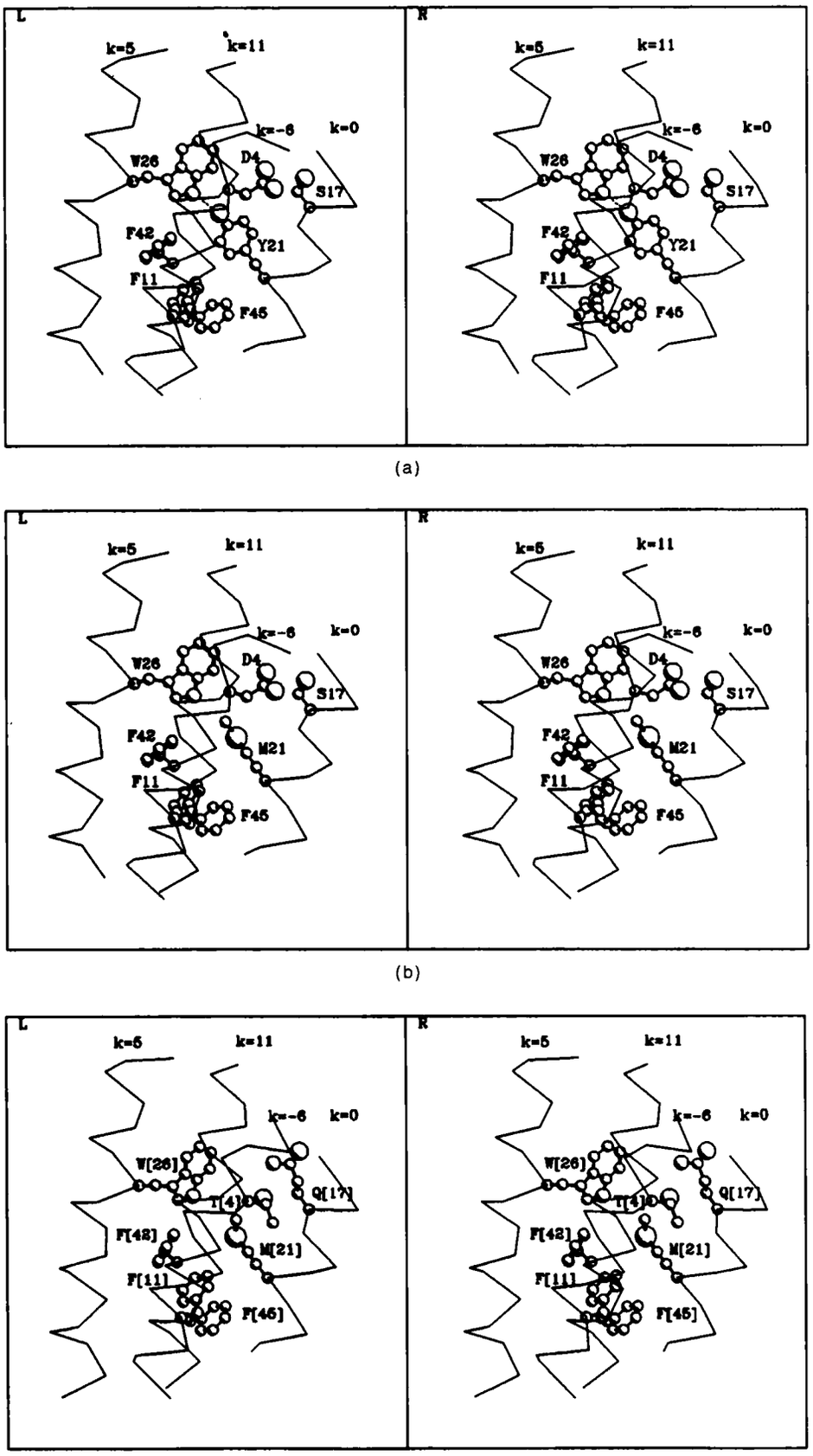


Figure 10. Environment of Trp in the model. Virion axis vertical. View from outside the virion along a radius (as Fig. 1(a)), showing central regions of units $k = 0$, 5 and 11; and the N-terminal region of $k = -6$. All side-chains except those labelled have been removed for clarity. The α-helix backbone is shown as a thin line connecting successive C^{α} atoms, and side-chains are shown as heavy lines, with atom types coded by size and shaded as in Fig. 1. (a) Wild-type fd^{C}, with hydrogen bond shown as a broken line between W26 and Y21. (b) Model of f1g8Y21M. (c) Wild-type If1.

Lys without steric problems, supporting the interpretation that the effect has to do with overall charge distribution on the inner surface of the protein shell, not detailed molecular structure.

(iii) *The interior of the protein shell*

Most of the residues involved in interactions between subunits in the interior of the protein shell are apolar and small. The degree to which a residue is buried in the interior can be measured by environmental analysis and Table 1 shows which residues are buried by this criterion. There are 13 sites with identical apolar residues in strains Ff, If1 and IKe (Table 1), and these are all either buried or partially buried. Residues Tyr21 and Phe11 are both buried by neighbouring subunits in the model of wild-type fd, and comparison with analogous class I strains suggests possible amino acid exchanges at these sites (Figs 9 and 10).

The three Phe residues are arranged in the Ff model so that Phe11 on one subunit intercalates between Phe42 and Phe45 on a neighbouring subunit (Fig. 9(a)). Laser Raman spectroscopy indicates less stacking of Phe rings at low ionic strength than at high ionic strength (Thomas *et al.*, 1983). This is consistent with a bending of the N-terminal region away from the body of the virion at low ionic strength, as discussed below to explain the response of Trp26 to ionic strength. In IKe, the residue at position 11 is Met, and this suggested that an F11M exchange in fd might be feasible. We have constructed this mutant and find that the F11M protein is expressed in infected bacteria and processed for membrane insertion, but no viable virions are produced. Figure 9(b) illustrates that the F11M exchange may be forbidden because Met11 in fd would interfere with Lys8, Tyr21 and Tyr24, whereas positions 8, 21 and 24 are all occupied by smaller residues in IKe (Fig. 9(c)). We then reasoned that the F11Y exchange in fd might be feasible, because Tyr could occupy the same intercalation site as Phe, and indeed fdg8F11Y is viable.

We have tried several mutagenesis experiments on f1 at sites in the apolar interior of the protein shell. Comparison of Ff with If1 at position 21 (Table 1; Fig. 10) suggests that a Y21M exchange might be feasible. Both Y21M and Y24M mutagenesis experiments were successful at a frequency of 30%. The X-ray diffraction pattern of the f1g8Y21M mutant at neutral pH (Fig. 4(b)) does not have the layer-line fanning observed for fd in Figure 4(a). The model of wild-type fd in the Tyr21 region is shown in Figure 10(a), and the model of Y21M is shown in Figure 10(b).

Numerous attempts to generate the A27P exchange in f1 were unsuccessful, even though IKe has Pro at this position (Table 1). Inspection of the models shows no obvious reason why Pro[27] should be permitted in IKe but forbidden in fd. However, there may be an indirect effect on Trp[26]. The disruption of the α-helix hydrogen-bonding pattern caused by inserting Pro (Barlow & Thornton, 1988) would change the orientation of the Trp[26] side-chain and thereby disrupt its hydrogen bond to Tyr[21] in fd (Fig. 10(a)); but in IKe position 21 is occupied by Leu, and no such hydrogen bond is present.

Non-specific mutagenesis using oligonucleotides that should replace Ile with Ile, Val, Leu or Phe gave eight Ile mutants (that is, with a mutant codon) and two Val exchanges at position 22 out of ten mutants analysed. The same kind of experiment at position 32 gave only Ile (mutant codon) out of two mutants analysed, even though position 32 is occupied by Val in both If1 and IKe (Table 1). Since Val is smaller than Ile, this exchange should be acceptable in the completed virion, and the failure of the mutation may reflect specific requirements during assembly.

The model shows that the Ff Trp26 is covered only by the N-terminal region of the $k = -11$ subunit (Fig. 10(a)), which may well bend away from the helix axis (perhaps using Pro6 as a hinge) at low ionic strength, when the negative charges on the surface of the virion become unshielded and repel each other. For strains If1 and IKe, the covering of Trp[26] by the N terminus is greater than for fd because of the additional residues at the N terminus, and Pro[6] is replaced (Fig. 10(c)). Both these facts suggest that for these two strains, the sensitivity of Trp[26] to ionic strength might be less than for fd. In both If1 and IKe, position 21 is occupied by an apolar residue, but a hydrogen bond from Trp[26] to Thr[4] might be possible (Fig. 10(c)), comparable to the Trp26 to Tyr21 hydrogen bond in Ff.

(iv) *The outer surface of the protein shell*

Titration of the intact fd virion over the range pH 3 to pH 9 is consistent with the location of the charged groups (NH_2 terminus, Glu2, Asp4, Asp5, Lys8, Asp12 and Glu20) on the exposed outer surface of the model (Zimmermann *et al.*, 1986), as illustrated in Figure 11(a). There is one basic residue on the outside surface of the class I model: Lys[8] in fd

Table 4
Amino acid residue exchanges in the Ff N-terminal region

	Residue position			
Strain	2	5	6	12
fd/f1	E−	D−	P	D−
M13	E−	D−	P	N
D53	E−	D−	S	N
*am*8H1R1	Q	D−	S	N
*am*8H1R2	S	D−	S	N
*am*8H1R5	Y	D−	S	D−
*am*8H1R6	L	D−	S	D−
R240(*su*-Ser)	S	D−	S	D−
R235	E−	H	P	D−
*am*8H1R268(*su*-Ser)	S	D−	P	D−
*am*8H1R265(*su*-Ser)	S	D−	F	N

Data are from Boeke *et al.* (1980) and Moses & Horiuchi (1982). The amber mutants were grown in suppressor strains that insert Ser (*su*-Ser) at the amber codon.

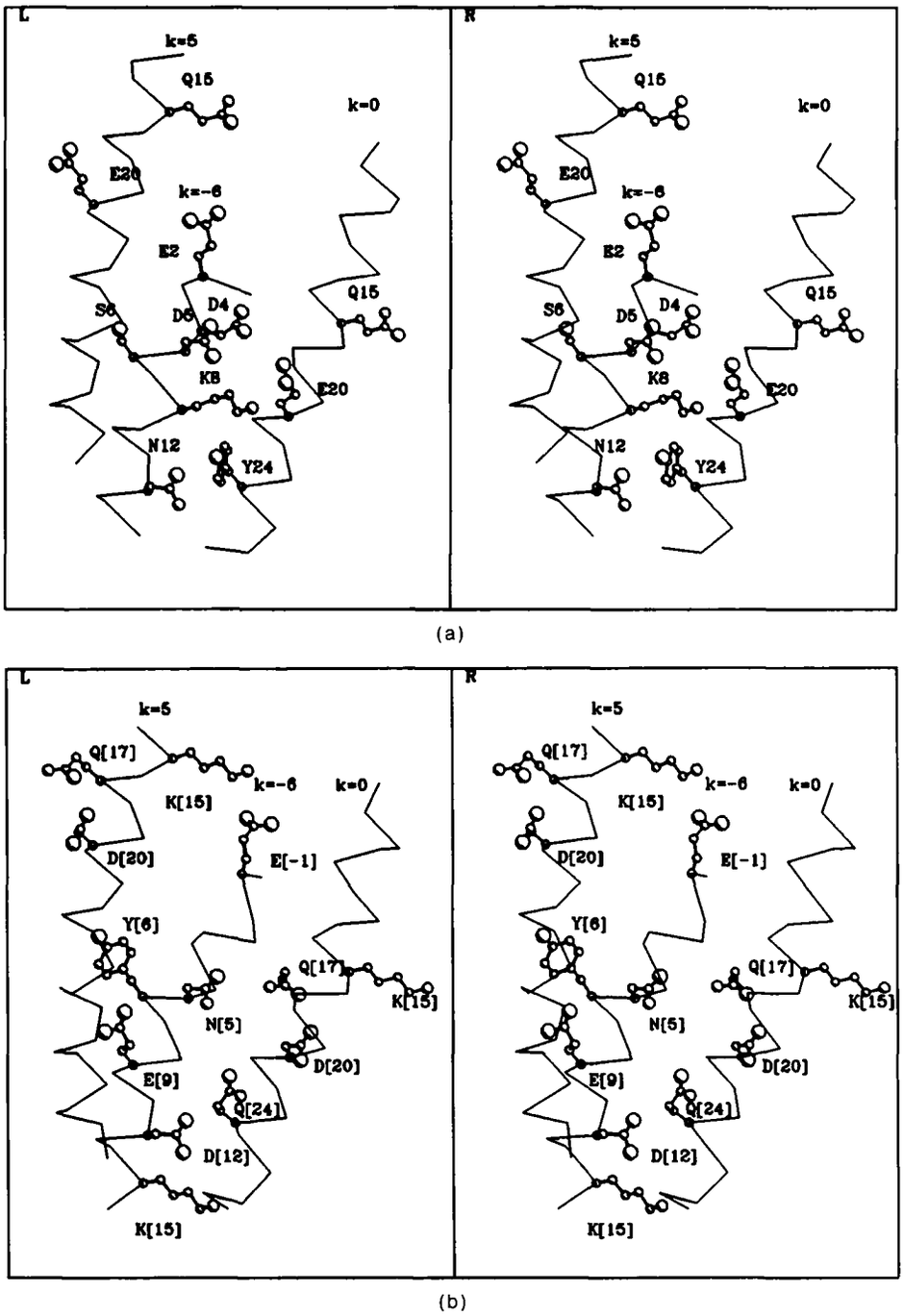


Figure 11. Side-chains on the outer surface of the virion. Virion axis vertical; view from outside the virion along a radius towards the virion axis (as Fig. 10), showing central regions of units $k = 0$ and 5, and the N-terminal region of $k = -6$. All side-chains except those labelled have been removed for clarity. The α-helix backbone is shown as a thin line connecting successive C^{α} atoms, and side-chains are shown as heavy lines, with atom types coded by size and shaded as in Fig. 1. (a) Charged groups on the outer surface of the fd^{C} model. Pro6 has been replaced by Ser and Asp12 by Asn in this Figure to illustrate the D53 mutant (see Table 4). (b) Charged groups on outer surface of Ike.

and If1, and Lys[15] in IKe (Table 1). The fd Lys8 is accessible to amidination without disrupting the virion, whereas the other four Lys in fd are not accessible (Armstrong *et al.*, 1983).

Salt bridges between residues i and $i + 4$ in an α-helix have been discussed as stabilizing features in the α-helix (Dao-pin *et al.*, 1991; Scholtz & Baldwin, 1992). Such salt bridges are feasible between Lys[8] and Asp[12] in fd and If1, and between Asp4 and Lys8 in fd (Fig. 11(a)). Extensive hydrogen-bond/salt bridge networks between, as well as within, subunits of both fd and IKe appear feasible from Figure 11. For example, in the fd and If1 models, −6Lys[8] can make a salt-bridge to Glu[20] (Fig. 11(a)). In IKe, 5Lys[15] can make a salt-bridge to −6Glu[−1], and Gln[24] can donate hydrogen

bonds to Asp[20] and to −6Asp[12] (Fig. 11(b)).

A single amber mutant in Ff gene *8* has been isolated (Pratt *et al.*, 1969), at position 2. This has enabled isolation of strains with various amino-acid residue exchanges at this position, either pseudo-revertants or suppressed amber mutants. Some of these replacement amino-acid residues are uncharged, so any requirement for an acidic residue in position 2 (see Discussion, below) cannot be stringent. Mutants have also been isolated with amino-acid exchanges at positions 5 (D5H), 6 (P6S or P6F) and 12 (D12N). Combinations of these exchanges that have been found in viable virions are listed in Table 4. Exchanges between charged and neutral side-chains have been identified at positions 2, 5 and 12, and these exchanges cause corresponding changes in the electrophoretic mobility of intact virions (Boeke *et al.*, 1980), supporting the view that these positions are on the outside surface of the virion (Fig. 11(a)). M13 wild-type g8p differs from fd by a single amino-acid residue exchange: Asp12 in fd becomes Asn12 in M13. Since Asp12 is accessible (Table 1) on the outside of the model, this exchange has little effect on the structure, but it would permit hydrogen-bonding from −6Asn12 to Tyr24 (Fig. 11(a)).

(d) *The ends of the virion*

At the ends of the virion, extensive apolar regions of the g8p subunit array would be exposed to solvent if it were not for the minor proteins that cap the two ends and probably stabilize the array by masking the apolar g8p sequences. Fibre diffraction studies of the virion do not give direct information about structure at the ends of the g8p array, but this structure can be explored by molecular modelling (Makowski, 1992). If the g8p subunit shape, symmetry, and nearest-neighbour contacts in the virion are maintained all the way to each end of the array, the N-terminal end of the array will have a hollow cup shape, and the C-terminal end will have a pointed arrowhead shape (Fig. 12).

To analyse the nature of the surface exposed at the two ends of the model, we use the methods of Eisenberg & McLachlan (1986). The net solvation free energy of the fd model, subdivided by atom type, is presented in Table 5. The subunit in the virion is stabilized by about 37 kcal/mol with respect to the isolated subunit, which is similar to the value 34 kcal/mol calculated using a slightly different set of atomic solvation parameters and a slightly different model (Marvin, 1990). The reduction of solvation free energy upon assembly comes entirely from masking apolar C. Masking other atoms which favour a polar environment reduces this value slightly. The model is stabilized by apolar interactions, as predicted from the resistance of the virion to extremes of salt and pH, and its sensitivity to apolar solvents. The stabilization of the subunit at the two ends of the array by the solvation free energy (Table 5) was calculated using accessibility of one subunit in contact with an appropriate subset of its neighbours. The g8p subunits at each end of the virion are stabilized by interaction with neighbouring subunits to about half the extent calculated for subunits within the virion array. Apolar g8p surfaces that are involved in contacts with other g8p subunits along the array are exposed at the two ends of the array, where they can interact with apolar surfaces of the minor proteins. The minor proteins g7p and g9p are at the end of the virion that is extruded first, and the g3p and g6p are at the end that is extruded last (Russel, 1991). It is often assumed that the N terminus of the g8p array is extruded first, and that therefore the initiator proteins g7p and g9p are associated with this end (Russel, 1991), but a model for initiation at the C terminus of the array has also been proposed (Marvin, 1989).

Residue accessibility (Table 6) at the N-terminal end of the array is as high as in the isolated subunit for residues near the N terminus of the g8p, and is the same as in the virion for residues near the C terminus of the g8p. The C-terminal end of the array shows a complementary pattern of access-

Table 5

Solvation free energy of fd[C]

Type	Net energy ΔG_S (kcal/mol)		
	Virion	N terminus	C terminus
Non-polar (C)	−48·5	−26·4	−25·5
Polar (neutral N/O)	3·4	2·0	1·7
Sulphur	0·2	0·2	0·2
Charged O^-	2·3	1·3	1·0
Charged N^+	6·0	3·3	4·8
$\Sigma\Delta G_S$	−36·6	−19·7	−17·9

Calculated as described by Eisenberg & McLachlan (1986), using the modified atomic solvation parameters of Eisenberg *et al.* (1989). The solvation free energy of the virion was calculated using accessibility of a subunit surrounded by nearest neighbours in the ±1, ±5, ±6, ±11 and ±17 directions. The accessibility at the N terminus of the virion (that is, the end of the virion towards which the N terminus of the subunit is directed), was calculated by surrounding the 0 subunit with only the subunits for $k = -17, -11, -6, -5, -1$ and 1. The accessibility at the C terminus of the virion was calculated using subunits $k = -1, 1, 5, 6, 11$ and 17. From the total solvation free energy for each type of array, the solvation free energy of a single isolated subunit was subtracted, to give the net solvation free energy ΔG_S of the array.

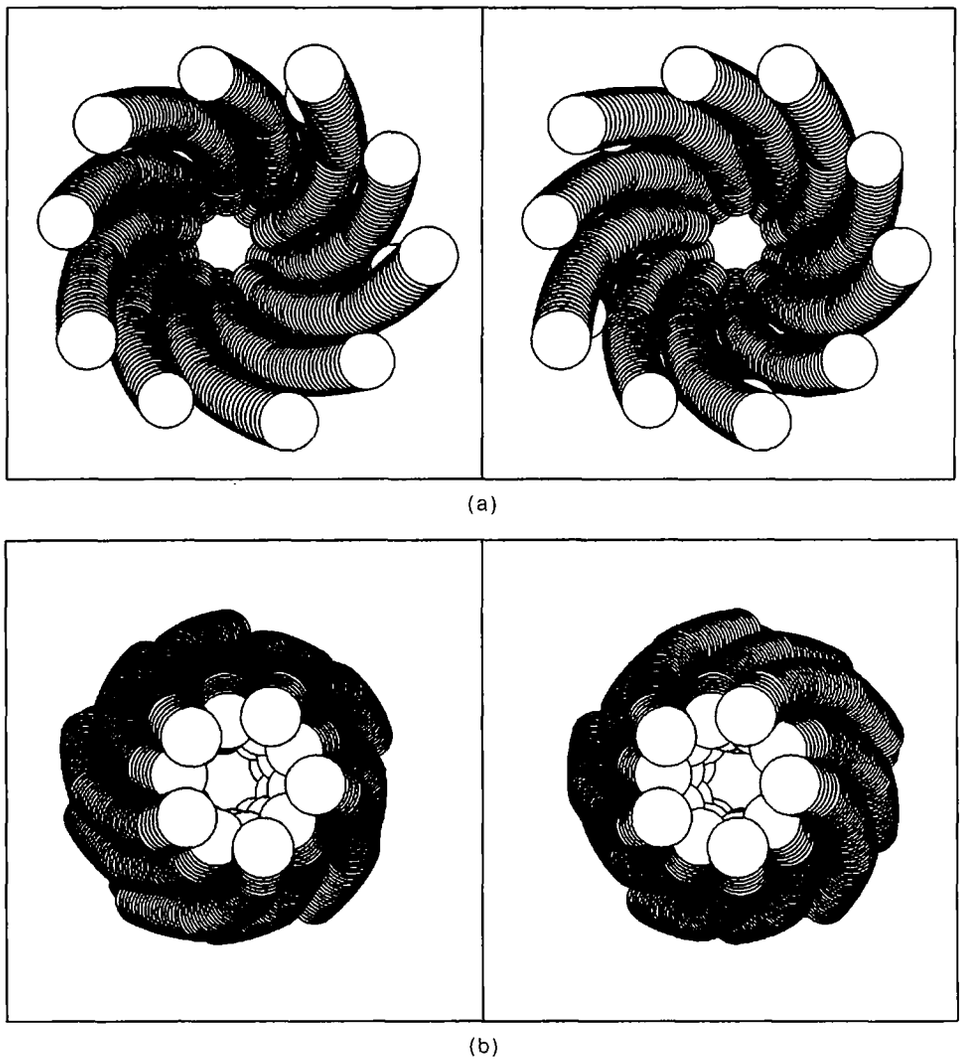


Figure 12. Ends of the fd virion. Each subunit is represented by a solid rod, as in Fig. 3(a). Symmetry-related copies of the subunit were generated using the virion helix parameters. View parallel to the virion helix axis. (a) View of the N-terminal end, towards decreasing unit number (decreasing z), units ..., −6, −5, −4, −3, −2, −1, 0. (b) View of the C-terminal end, towards increasing unit number (increasing z), units 0, 1, 2, 3, 4, 5, 6, ... Units 1 to 4 are the same as units −4 to −1 (Marvin, 1989).

ibility, although the C terminus of the g8p is not completely accessible even here, because symmetry-related units ±1 make close contacts with unit 0 at the C terminus (Fig. 8) but not at the N terminus of g8p (Fig. 12).

4. Discussion

Laser Raman spectroscopy of wild-type class I virions indicates that Tyr21 and Tyr24 of fd and Tyr[24] of If1 accept hydrogen bonds from positive donor groups (Thomas *et al.*, 1983). The evidence that Trp26 may donate a hydrogen bond to −5Tyr21 is discussed below and illustrated in Figure 10(a). Residue Tyr[24] of both fd and If1 can accept a hydrogen bond from Lys[8], as shown in Figure 11(a). These experiments on wild-type fd are further supported by spectroscopy of the Y21M and Y24M mutants (G. J. Thomas Jr, personal communication). The spectrum due to Tyr24, determined by measuring the Y21M mutant, is consistent with Tyr24 being both a donor and an acceptor of hydrogen bonding, and is therefore consistent with the position of Tyr24 on the outside surface of the model, where it can donate a hydrogen bond to water, but is near enough to Lys8 to accept a hydrogen bond. The spectrum due to Tyr21 was determined in two ways: by taking the difference between the spectrum of the Y21M mutant and the spectrum of wild-type f1, and assuming this to be the spectrum due to Tyr21; and by measuring the spectrum of the Y24M mutant. Both measurements are consistent with Tyr21 being a strong hydrogen-bond acceptor, and are consistent with the proposal that Trp26 donates a hydrogen bond to Tyr21. Residue Tyr[6] in IKe is both a hydrogen-bond donor and acceptor, and is therefore probably accessible to solvent (Thomas *et al.*, 1983), consistent with the partial accessibility calculated for position 6 (Table 1 and Fig. 11(b)).

The Trp26 conformation has also been studied by laser Raman spectroscopy (Thomas *et al.*, 1983;

Table 6
Environment and accessibility at the ends of fd[D]

	Environment†		Accessibility‡		
Position	N-term	C-term	Centre	N-term	C-term
1	E	E	0·9	1·0	0·8
2	E	E	1·0(1·0)	1·0(1·0)	1·0(1·0)
3	E	E	—	—	—
4	E	P1	0·1(0·3)	1·0(1·0)	0·1(0·3)
5	E	E	1·0(1·0)	1·0(1·0)	1·0(1·0)
6	E	P2	0·9	1·0	0·9
7	E	P1	0·0	1·0	0·0
8	P2	B3	0·4(0·7)	1·0(1·0)	0·4(0·6)
9	E	E	1·0	1·0	1·0
10	E	P1	0·2	1·0	0·2
11	E	B1	0·0	1·0	0·0
12	E	P2	0·7(0·7)	1·0(1·0)	0·7(0·7)
13	E	E	0·9	0·9	1·0
14	P2	B2	0·2	1·0	0·2
15	E	B2	0·2	1·0	0·2
16	E	E	1·0	1·0	1·0
17	P2	P2	0·2	0·3	1·0
18	E	P1	0·0	1·0	0·0
19	E	P2	0·8	0·9	0·8
20	P2	P2	0·6(0·5)	0·6(0·6)	1·0(1·0)
21	B2	B2	0·0	0·4	0·3
22	P2	B1	0·0	1·0	0·0
23	E	E	—	—	—
24	B3	P2	0·3	0·3	0·9
25	E	P1	0·1	1·0	0·1
26	B3	B3	0·1	0·5	0·5
27	E	E	0·7	0·7	1·0
28	P1	B2	0·0	0·5	0·3
29	P2	B1	0·0	1·0	0·0
30	P1	P2	0·4	0·4	0·8
31	P1	E	0·3	0·3	1·0
32	P1	B1	0·1	0·9	0·2
33	P1	B2	0·0	0·6	0·1
34	E	E	—	—	—
35	P1	E	0·1	0·2	0·9
36	E	P1	0·1	1·0	0·0
37	B1	P2	0·0	0·1	0·8
38	E	E	—	—	—
39	B2	P1	0·0	0·3	0·6
40	P2	B3	0·3(0·7)	1·0(1·0)	0·3(0·6)
41	B1	P2	0·0	0·1	1·0
42	B1	E	0·0	0·0	1·0
43	B3	B3	0·2(0·1)	0·5(0·8)	0·4(0·1)
44	B2	B3	0·3(0·4)	0·4(0·4)	0·5(0·4)
45	B1	E	0·0	0·0	1·0
46	P2	E	0·3	0·1	0·7
47	E	E	1·0	1·0	1·0
48	B1	P2	0·1(0·1)	0·1(0·3)	0·8(0·4)
49	P1	E	0·0	0·0	1·0
50	P2	E	0·6	0·6	1·0

†Environment calculations are described in Table 1. N-term, refers to the environment at the N terminus of the virion (that is, the end of the virion towards which the N terminus of the subunit is directed); C-term, refers to the environment at the C terminus of the virion, calculated using subsets of the neighbouring subunits as described in the legend to Table 5.

‡Accessibility of fd side-chains was calculated as described by Richards (1985), using a rolling sphere of 1·4 Å radius, for side-chains on the subunit surrounded by its nearest neighbours in the assembly, and on the isolated single subunit. The Ala1 accessible area includes N, and the Ser50 area includes COO. The relative accessibility (ratio of accessibility in the assembly to accessibility in the isolated subunit) is tabulated. Numbers in parentheses are the relative accessibilities of charged atoms ($O^{\epsilon 2}$ of Glu, $O^{\delta 2}$ of Asp and N^{ζ} of Lys). Centre, is accessibility along the shaft of the virion, corresponding to the environment calculations of Table 1; N-term, is accessibility at the N terminus of the virion; C-term, is accessibility at the C terminus of the virion, corresponding to the environment calculations in this Table.

Aubrey & Thomas, 1991). The indole 1-NH donor group forms a hydrogen bond, probably to a hydroxyl oxygen acceptor (Aubrey & Thomas, 1991). In our refined fd^C model, the hydroxyl O of −5Tyr21 in the model is 2·73 Å from the indole 1-NH of Trp26, a good hydrogen-bonding distance (Fig. 10(a)).

Flow linear dichroism spectra of fd and IKe indicate that the angle between the plane of the Trp[26] side-chain and the virion axis is no more than 43° for fd and 41° for IKe (Clack & Gray, 1992). To determine this angle for our models, we calculate the direction cosines of the line normal to the least-squares plane through the Trp ring, and take the angle that this normal makes with the virion axis as the complement of the angle between the plane and the virion axis. For our fd^C model, the angle between the plane of the Trp and the virion axis is 10°, increasing to 16° for fd^D, and for our IKe model it is 37°.

Oxidation of Trp26 with *N*-bromosuccinimide is accompanied by changes in spectra attributed to electronic coupling between Trp and Phe (Arnold *et al.*, 1992). Trp26 and Phe42 are in van der Waals contact in the model (Fig. 10(a)), with 3·5 Å between $C^{\delta 1}$ of Trp26 and $C^{\epsilon 1}$ of 6Phe42. The spectroscopic and oxidation experiments indicate that the Trp26 is partially buried and immobilized, and yet is accessible to solvent. A flexible N-terminal region would be consistent with these observations.

If Pro6 is replaced by Ser in M13 (strain D53 in Table 4 and Fig. 11(a)), the buoyant density of the virion in a CsCl gradient is increased by about 0·01 g/ml, or about 0·8%, and other mutants in g8p also have altered buoyant density (Pratt *et al.*, 1969). The Pro6 is accessible (Table 1), and in the P6S model, the Ser oxygen atom is also accessible to solvent (Fig. 11(a)). Solvent interacts more closely with polar Ser than with apolar Pro, so the excluded volume of the P6S virion is expected to be less and its buoyant density greater than wild-type M13. Quantitative calculations using the residue volumes tabulated by Nave *et al.* (1981) show that the increase in partial specific volume expected for the P6S mutant is about 0·3%, consistent with the observed change in buoyant density. Replacing Glu2 by Tyr or Leu in M13P6S is always accompanied by a further N12D mutation (Table 4). There is no obvious reason why E2L/P6S and E2Y/P6S should be forbidden in the M13 structure, but the mutations may interfere with processing of the leader sequence on the precursor protein (Boeke *et al.*, 1980). Virions can be produced from bacteria that are expressing both wild-type f1 and P6S/D12N, E2S/P6S/D12N or E2S/P6S coat proteins, and these virions have mixtures of the two proteins (Moses & Horiuchi, 1982).

The interlocking of side-chains on neighbouring subunits is a direct consequence of close-packed nearly parallel α-helices (Chothia *et al.*, 1981). Slight changes in length of the virion axial repeat with hydration (Dunker *et al.*, 1974) reflect the fact that the interlocking of side-chains is slightly flexible, but in the completed virion, the pattern of interlocking cannot change. Before assembly of the virion, neighbouring α-helix subunits in the membrane are thought to pre-assemble as uniform rods, but they cannot have the same side-chain interlocking as in the finished virion. During assembly, neighbouring α-helices must therefore slide along their length with respect to one another, until the final local side-chain packing is reached. This sliding must be relatively independent of specific side-chains, but it is possible that certain side-chain replacements could interfere with the sliding mechanism and thereby be lethal. The local packing of α-helix rods in the membrane pre-assembly may be geometrically related to the final symmetry of the virion, or it may be unspecific (Marvin & Wachtel, 1976; Marvin, 1989; Dunker *et al.*, 1991): this question remains to be answered.

The protein model that we use here is 100% α-helix, consistent with some spectroscopic experiments (Clack & Gray, 1989), but some variations in the secondary structure may appear as further evidence is accumulated. The last few residues at either end might have i to $i + 3$ or $i + 5$ rather than α-helix (i to $i + 4$) hydrogen-bonding patterns, as sometimes found in globular proteins (Baker & Hubbard, 1984). The first few residues at the N terminus may be non-helical (Opella *et al.*, 1987). There must be some disruption of α-helix hydrogen-bonding in Ff at Pro6. The suggestion that some ionic-strength dependent spectroscopic properties of the virion are caused by bending of the α-helix at Pro6 could be tested by comparative studies on mutants with Pro6 replaced by Ser. Electrostatic interactions between the N terminus of an α-helix and an acidic residue second along the helix have been discussed by Richardson & Richardson (1988) and Nicholson *et al.* (1991). The acidic residue does not just interact with the charge on the N terminus of the polypeptide chain, because the effect is found for α-helix segments embedded within the polypeptide chain. Instead, the acidic residue is thought to be involved in compensating the charge of the α-helix dipole. The second residue of g8p in all three wild-type class I virions is Glu or Asp (Table 1).

Our model has been refined against quantitative low-resolution X-ray data and is supported by qualitative comparison with higher resolution data. The model is also consistent with spectroscopic experiments, chemical modification experiments, comparison of related viruses, and the effects of side-chain mutation. Although some details may change after additional refinement, especially at the two ends of the subunit, the model presented here forms a useful basis for design and interpretation of further spectroscopic, chemical modification and genetic manipulation experiments, and for exploring detailed proposals about viral assembly.

We are grateful to Professors R. N. Perham and G. Cesareni for advice and support; to Professor G. J. Thomas Jr for discussion of his laser Raman experiments; and to the staff of the Medical Research Council

Laboratory of Molecular Biology, Cambridge for advice and discussion and for generous access to computing and computer graphics facilities. The research in England was supported by grants from the Leverhulme Trust and the Science and Engineering Research Council; the research in Rome was supported by Target Project Ingegneria Genetica from CNR.

References

Armstrong, J., Hewitt, J. A. & Perham, R. N. (1983). Chemical modification of the coat protein in bacteriophage fd and orientation of the virion during assembly and disassembly. *EMBO J.* **2**, 1641–1646.

Arnold, G. E., Day, L. A. & Dunker, A. K. (1992). Tryptophan contributions to the unusual circular dichroism of fd bacteriophage. *Biochemistry*, **31**, 7948–7956.

Aubrey, K. L. & Thomas, G. J., Jr (1991). Raman spectroscopy of filamentous bacteriophage Ff (fd, M13, f1) incorporating specifically-deuterated alanine and tryptophan side chains. Assignments and structural interpretation. *Biophys. J.* **60**, 1337–1349.

Baker, E. N. & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Progr. Biophys. Mol. Biol.* **44**, 97–179.

Banner, D. W., Nave, C. & Marvin, D. A. (1981). Structure of the protein and DNA in fd filamentous bacterial virus. *Nature (London)*, **289**, 814–816.

Barlow, D. J. & Thornton, J. M. (1988). Helix geometry in proteins. *J. Mol. Biol.* **201**, 601–619.

Bernstein, F. C. Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Bhattacharjee, S., Glucksman, M. J. & Makowski, L. (1992). Structural polymorphism correlated to surface charge in filamentous bacteriophages. *Biophys. J.* **61**, 725–735.

Boeke, J. D., Russel, M. & Model, P. (1980). Processing of filamentous phage pre-coat protein. Effect of sequence variations near the signal peptidase cleavage site. *J. Mol. Biol.* **143**, 103–116.

Booy, F. P. & Fowler, A. G. (1985). Cryo-electron microscopy reveals macromolecular organization within biological liquid crystals seen in the polarizing microscope. *Int. J. Biol. Macromol.* **7**, 327–335.

Bowie, J. U., Lüthy, R. & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, **253**, 164–170.

Bryan, R. K. (1987). Maximum entropy in structural molecular biology: the fiber diffraction phase problem. In *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems* (Smith, C. R. & Erickson, G. J., eds), pp. 207–228, D. Reidel Publishing Company, Dordrecht.

Bryan, R. K., Bansal, M., Folkhard, W., Nave, C. & Marvin, D. A. (1983). Maximum-entropy calculation of the electron density at 4 Å resolution of Pf1 filamentous bacteriophage. *Proc. Nat. Acad. Sci., U.S.A.* **80**, 4728–4731.

Chothia, C., Levitt, M. & Richardson, D. (1981). Helix to helix packing in proteins. *J. Mol. Biol.* **145**, 215–250.

Clack, B. A. & Gray, D. M. (1989). A CD determination of the α-helix contents of the coat proteins of four filamentous bacteriophages: fd, IKe, Pf1 and Pf3. *Biopolymers*, **28**, 1861–1873.

Clack, B. A. & Gray, D. M. (1992). Flow linear dichroism spectra of four filamentous bacteriophages: DNA and coat protein contributions. *Biopolymers*, **32**, 795–810.

Cross, T. A. & Opella, S. J. (1985). Protein structure by solid state nuclear magnetic resonance. Residues 40 to 45 of bacteriophage fd coat protein. *J. Mol. Biol.* **182**, 367–381.

Dao-pin, S., Sauer, U., Nicholson, H. & Matthews, B. W. (1991). Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry*, **30**, 7142–7153.

Day, L. A. & Wiseman, R. L. (1978). A comparison of DNA packaging in the virions of fd, Xf and Pf1. In *The Single-Stranded DNA Phages* (Denhardt, D. T., Dressler, D. & Ray, D. S., eds), pp. 605–625, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Dunker, A. K., Klausner, R. D., Marvin, D. A. & Wiseman, R. L. (1974). Filamentous bacterial viruses. X. X-ray diffraction studies of the R4 A-protein mutant. *J. Mol. Biol.* **82**, 115–117.

Dunker, A. K., Ensign, L. D., Arnold, G. E. & Roberts, L. M. (1991). A model for fd phage penetration and assembly. *FEBS Letters*, **292**, 271–274.

Eisenberg, D. & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature (London)*, **319**, 199–203.

Eisenberg, D., Wesson, M. & Yamashita, M. (1989). Interpretation of protein folding and binding with atomic solvation parameters. *Chem. Scr.* **29A**, 217–221.

Fraser, R. D. B., MacRae, T. P., Miller, A. & Suzuki, E. (1983). Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **167**, 497–521.

Glucksman, M. J., Bhattacharjee, S. & Makowski, L. (1992). Three-dimensional structure of a cloning vector. X-ray diffraction studies of filamentous bacteriophage M13 at 7 Å resolution. *J. Mol. Biol.* **226**, 455–470.

Gray, T. M. & Matthews, B. W. (1984). Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.* **175**, 75–81.

Greenwood, J., Hunter, G. J. & Perham, R. N. (1991). Regulation of filamentous bacteriophage length by modification of electrostatic interactions between coat protein and DNA. *J. Mol. Biol.* **217**, 223–227.

Hunter, G. J., Rowitch, D. H. & Perham, R. N. (1987). Interactions between DNA and coat protein in the structure and assembly of filamentous bacteriophage fd. *Nature (London)*, **327**, 252–254.

Jack, A. & Levitt, M. (1978). Refinement of large structures by simultaneous minimization of energy and *R* factor. *Acta Crystallogr. sect. A*, **34**, 931–935.

Levitt, M. (1974). Energy refinement of hen egg-white lysozyme. *J. Mol. Biol.* **82**, 393–420.

Levitt, M. (1983). Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* **168**, 595–620.

Lüthy, R., Bowie, J. U. & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature (London)*, **356**, 83–85.

Makowski, L. (1992). Terminating a macromolecular helix. Structural model for the minor proteins of bacteriophage M13. *J. Mol. Biol.* **228**, 885–892.

Makowski, L. & Caspar, D. L. D. (1978). Filamentous bacteriophage Pf1 has 27 subunits in its axial repeat. In *The Single-Stranded DNA Phages* (Denhardt, D. T., Dressler, D. & Ray, D. S., eds), pp. 627–643,

Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Makowski, L. & Caspar, D. L. D. (1981). The symmetries of filamentous phage particles. *J. Mol. Biol.* **145**, 611–617.

Maret, G. & Dransfeld, K. (1985). Biomolecules and polymers in high steady magnetic fields. *Topics Appl. Phys.* **57**, 143–204.

Marvin, D. A. (1989). Dynamics of telescoping *Inovirus*: a mechanism for assembly at membrane adhesions. *Int. J. Biol. Macromol.* **11**, 159–164.

Marvin, D. A. (1990). Model-building studies of *Inovirus*: genetic variations on a geometric theme. *Int. J. Biol. Macromol.* **12**, 125–138.

Marvin, D. A. & Wachtel, E. J. (1975). Structure and assembly of filamentous bacterial viruses. *Nature (London)*, **253**, 19–23.

Marvin, D. A. & Wachtel, E. J. (1976). Structure and assembly of filamentous bacterial viruses. *Phil. Trans. Roy. Soc. ser. B*, **276**, 81–98.

Marvin, D. A. & Nave, C. (1982). X-ray fiber diffraction. In *Structural Molecular Biology* (Davies, D. B., Saenger, W. & Danyluk, S. S., eds), pp. 3–44, Plenum Publishing Corporation, New York.

Marvin, D. A., Pigram, W. J., Wiseman, R. L., Wachtel, E. J. & Marvin, F. J. (1974*a*). Filamentous bacterial viruses. XII. Molecular architecture of the class I (fd, If1, IKe) virion. *J. Mol. Biol* **88**, 581–600.

Marvin, D. A., Wiseman, R. L. & Wachtel, E. J. (1974*b*). Filamentous bacterial viruses. XI. Molecular architecture of the class II (Pf1, Xf) virion. *J. Mol. Biol.* **82**, 121–138.

Marvin, D. A., Bryan, R. K. & Nave, C. (1987). Pf1 *Inovirus*. Electron density distribution calculated by a maximum entropy algorithm from native fibre diffraction data to 3 Å resolution and single isomorphous replacement data to 5 Å resolution. *J. Mol. Biol.* **193**, 315–343.

Marvin, D. A., Nave, C., Bansal, M., Hale, R. D. & Salje, E. K. H. (1992). Two forms of Pf1 *Inovirus*: X-ray diffraction studies on a structural phase transition and a calculated libration normal mode of the asymmetric unit. *Phase Transit.* **39**, 45–80.

McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987). Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**, 295–310.

McLachlan, A. D. (1982). Rapid comparison of protein structures. *Acta Crystallogr. sect. A*, **38**, 871–873.

Model, P. & Russel, M. (1988). Filamentous bacteriophage. In *The Bacteriophages* (Calendar, R., ed.), vol. 2, pp. 375–456, Plenum Press, New York.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins Struct. Funct. Genet.* **12**, 345–364.

Moses, P. B. & Horiuchi, K. (1982). Effects of transposition and deletion upon coat protein gene expression in bacteriophage f1. *Virology*, **119**, 231–244.

Nakashima, Y., Wiseman, R. L., Konigsberg, W. & Marvin, D. A. (1975). Primary structure and sidechain interactions of Pf1 filamentous bacterial virus coat protein. *Nature (London)*, **253**, 68–71.

Nave, C., Fowler, A. G., Malsey, S., Marvin, D. A., Siegrist, H. & Wachtel, E. J. (1979). Macromolecular structural transitions in Pf1 filamentous bacterial virus. *Nature (London)*, **281**, 232–234.

Nave, C., Brown, R. S., Fowler, A. G., Ladner, J. E., Marvin, D. A., Provencher, S. W., Tsugita, A., Armstrong, J. & Perham, R. N. (1981). Pf1 filamentous bacterial virus. X-ray fibre diffraction analysis of two heavy-atom derivatives. *J. Mol. Biol.* **149**, 675–707.

Nicholson, H., Anderson, D. E., Dao-pin, S. & Matthews, B. W. (1991). Analysis of the interaction between charged side chains and the α-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry*, **30**, 9816–9828.

Opella, S. J., Stewart, P. L. & Valentine, K. G. (1987). Protein structure by solid-state NMR spectroscopy. *Quart. Rev. Biophys.* **19**, 7–49.

Ponder, J. W. & Richards, F. M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791.

Pratt, D., Tzagoloff, H. & Beaudoin, J. (1969). Conditional lethal mutants of the small filamentous coliphage M13. *Virology*, **39**, 42–53.

Provencher, S. W. & Glöckner, J. (1982). Rapid analytical approximations for disorientation integrals in fiber diffraction. *J. Appl. Crystallogr.* **15**, 132–135.

Rasched, I. & Oberer, E. (1986). Ff coliphages: structural and functional relationships. *Microbiol. Rev.* **50**, 401–427.

Richards, F. M. (1985). Calculation of molecular volumes and areas for structures of known geometry. *Methods Enzymol.* **115B**, 440–464.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α-helices. *Science*, **240**, 1648–1652.

Russel, M. (1991). Filamentous phage assembly. *Mol. Microbiol.* **5**, 1607–1613.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scholtz, J. M. & Baldwin, R. L. (1992). The mechanism of α-helix formation by peptides. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 95–118.

Summers, N. L., Carlson, W. D. & Karplus, M. (1987). Analysis of side-chain orientations in homologous proteins. *J. Mol. Biol.* **196**, 175–198.

Thomas, G. J., Jr, Prescott, B. & Day, L. A. (1983). Structure similarity, difference and variability in the filamentous viruses fd, If1, IKe, Pf1 and Xf. Investigation by laser Raman spectroscopy. *J. Mol. Biol.* **165**, 321–356.

Torbet, J. & Maret, G. (1979). Fibres of highly oriented Pf1 bacteriophage produced in a strong magnetic field. *J. Mol. Biol.* **134**, 843–845.

Wachtel, E. J., Wiseman, R. L., Pigram, W. J., Marvin, D. A. & Manuelidis, L. (1974). Filamentous bacterial viruses. XIII. Molecular structure of the virion in projection. *J. Mol. Biol.* **88**, 601–618.

Zimmermann, K., Hagedorn, H., Heuck, C. C., Hinrichsen, M. & Ludwig, H. (1986). The ionic properties of the filamentous bacteriophages Pf1 and fd. *J. Biol. Chem.* **261**, 1653–1655.

Edited by D. DeRosier

(Received 21 September 1992; accepted 31 August 1993)