DNA AND PROTEIN LATTICE-LATTICE INTERACTIONS
IN THE FILAMENTOUS BACTERIOPHAGES

CHRISTOPHER J. MARZEC AND LOREN A. DAY
The Public Health Research Institute of the City of New York, New York, New York, 10016

ABSTRACT   The relations between the protein coats and DNAs of the four filamentous bacteriophages fd, Xf, Pf1, and Pf3 are considered. These viruses have similar morphologies, yet show a diversity of detailed structure, having different protein coat symmetries (helical and rotational), different coat protein sizes (44–50 amino acids per subunit) and sequences, different nucleotide axial translations (2.3–5.5 Å), and different ratios of nucleotides per coat protein subunit (integers 1.0 and 2.0, and nonintegers ~2.4). These divergences are all reconciled quantitatively by means of two theoretical concepts: the pitch connection and the restricted pitch connection. The pitch connection relates protein and DNA surface lattices with arbitrary, nonintegral nucelotide/subunit ratios in a nonrandom way. The restricted pitch connection selects a preferred set of n/s values. Both relations are derived formally in a mathematical appendix. The available structural data are explained, including the fd DNA pitch indicated by x-ray diffraction photos and the similar DNA morphologies of Xf and fd. Predictions are made for the existence of nonclassical inverted DNA structures (I-DNA) in Pf1 and Pf3.

INTRODUCTION

The filamentous phages (e.g., Pf1, Xf, Pf3, fd, Ike, If1) contain a circular single-stranded genome ensheathed by a protein coat consisting of several thousand copies of a major coat protein and a few copies each of one or more minor coat proteins. In each of these viruses the topologically circular DNA is packaged as two antiparallel chains occupying a central core region. These are not Watson-Crick paired throughout the length of the virion, although a small amount of Watson-Crick pairing may be present in some structures. Table I states physical parameters for Pf1, Xf, Pf3, and fd, the most significant parameters among them for this paper being the nucleotide/subunit ratios (n/s). Two of these viruses (Pf1 and Xf) have what appear to be integer values and two viruses (fd and Pf3) definitely have noninteger values. This article is concerned with two questions involving the DNA-protein interfaces of the filamentous viruses: (a) What manner of protein-nucleotide interaction allows a noninteger value of n/s and a matching of the different protein and DNA symmetries? and (b) Can particular values of n/s be interpreted in terms of some strategy for producing an optimum structure?

In order even to pose these questions, we have adopted the hypothesis that the DNA-protein interface is well-ordered and understandable through a simple model. This is not the only viewpoint. It has been asserted that in fd the DNA-protein interaction is "non-uniform" (Banner et al., 1981), and also that "the helical symmetry of the DNA and the coat protein must be unrelated" (Makowski and

Caspar, 1981). Part of the issue is semantic. The noninte-gral n/s ratios in fd and Pf3 are themselves sufficient to imply "non-uniformity," if a uniform interaction is one in which each major coat-protein subunit has the same nucleotide environment. Similarly, one might say that two coaxial helices are "unrelated" if their pitches differ, however else they may interact. In any case, some degree of structural interdependency for the DNA and protein sheaths is suggested by the fact that for none of these viruses has a protein sheath without DNA (ghost) or a DNA molecule in its native configuration, but without protein, been observed; but this may result from an assembly mechanism. An infrared linear dichroism study (Fritzsche et al., 1981) finds that the fd "structure involves a rigidly held DNA molecule complexed within a stable coat protein matrix." Also, electron micrographs and sedimentation analysis of sheared f1 (fd) virions show that the DNA does not unravel from the cut ends (Webster et. al., 1981; Grant et al., 1981), indicating that it is not simply stuffed into the protein coat but is somehow bound. Furthermore, Ag^+ complexes with Pf3 (private communication with A. Casadevall), Xf, and fd (Casadevall and Day, 1982) show increased sedimentation coefficients, compared with the native virions. The magnitudes of the increases are much larger than can be accounted for by the addition of the Ag^+ mass alone, so the changes are due to overall perturbations in the structures or in the hydrations of these virions (Casadevall and Day, 1982). Because the Ag^+ binds to the DNA and not to the protein, the DNA perturbation induced by the Ag^+ must propagate outward into the protein, implying a strong coupling in these virions

	Pf1	Xf	Pf3	fd
Length*	2.0 μm	1.0 μm	0.7 μm	0.9 μm
Nucleotides per Subunit‡	0.97 ± 0.05	2.02 ± 0.04	2.38 ± 0.14	2.32 ± 0.08
Symmetry of protein sheath§	27u/5t	27u/5t	27u/5t	5-start
Axial nucleotide separation (one chain)‖	5.73 Å	2.75 Å	2.33 Å	2.76 Å

Properties common to all four viruses are mass per length, which is in the vicinity of 1,800–1,900 daltons/Å for each (Newman et al., 1977; Berkowitz and Day, 1980; Nave et al., 1981; Newman et al., 1982), and the small size, from 44–50 amino acids, and relatively high alpha-helicity, from 50 to 100%, of the major coat protein subunits (data compiled in Day and Wiseman, 1978; for Pf3, Thomas and Day, 1981, and D. Gluck Putterman and B. Frangione, personal communications).

*‡Data compiled and discussed by Day and Wiseman (1978) and, for Pf3, by Newman et al. (1982).

§Pf1, Xf, and Pf3 all have 27 subunits in five turns of the protein helix over repeat distances of 75 Å (Marvin et al., 1974; Wiseman and Day, 1977; Nave et al., 1981; for Pf3, C. Peterson et al. (1982); fd has five subunits every 16 Å (Newman et al., 1977) arranged with fivefold rotational symmetry (Makowski and Caspar, 1978, 1981).

‖Values calculated from x-ray unit cell dimensions and the data of ‡ and §. The values are accurate to 6% or better. Note that the best values generated by the theoretical treatment of this paper are (‡) 1, 2, 2.4, and 2.4; and (‖) 5.55, 2.77, 2.31, and 2.67 Å.

between the DNA and protein sheath. Finally, x-ray diffraction photos of magnetically oriented fd fibers show layer lines ascribed to the DNA (Banner et al., 1981), which requires regular DNA helices in fd. We will show both that a nonuniform nucleotide environment does not imply that the DNA-protein interaction is random and that, although different, the DNA and protein symmetries in fd and the other filamentous viruses can be neatly connected.

We will answer the first of our questions above by means of a geometrical construction that we call the "pitch connection" and the second by means of the "restricted pitch connection." We provide here some context for these concepts by observing that in the filamentous phages, and in rodlike or filamentous nucleoproteins in general, the nucleotide-protein interactions can be described in order of decreasing symmetry: (a) "simple periodic" in which every nth nucleotide experiences the same protein environment (n = 3 for TMV), and the protein and the nucleic acid have the same helical symmetry; (b) "restrictively pitch connected," in which the protein and nucleic acid symmetries are different but related by the pitch connection, subject to the condition that every nth nucleotide of one strand experiences the same protein environment; (c) "simple pitch connected," in which there is not necessarily a

periodicity of nucleotide environments, but the nucleic acid and protein symmetries are related at the level of secondary structure; (d) "irregular," in which the nucleic acid and protein coat are unrelated.

THE PITCH CONNECTION

The pitch connection (Marzec and Day, 1980) affords a simple and flexible answer to the symmetry-matching problems posed by the occurrence of nonintegral n/s values in both fd and Pf3 and by the occurrence of helical DNA symmetry and rotational protein-coat symmetry in fd.

Fig. 1 a shows a helical DNA surface lattice not intended to represent any known structure, but drawn solely to demonstrate lattice interactions. In this and subsequent lattice diagrams, we adopt the convention that the lattice is drawn as it would be seen from outside the virus, so a right-handed helix is represented as a straight line with a positive slope. Each nucleotide (X for the up strand and 0 for the down strand) is positioned at its z and θ polar coordinates, with z directed along the virion structure axis, and coordinates $\theta = 0$ and $\theta = 2\pi$ identical. The solid line represents the helix traversed by the sugar-phosphate backbone, with pitch P_{DNA}. An X or 0 represents the point of a nucleotide that lies farthest from the structure axis. In Fig. 1b the backbone helix has been omitted. The nucleotides have been joined instead by dashed lines that represent the ridges formed by the outermost segments of each nucleotide. The ridge lines could have been drawn differently in Fig. 1 b to demonstrate different possibilities consistent with the same underlying nucleotide surface lattice. Such constructions are reminiscent of those used by Chothia et al. (1977) for analyzing α-helix packing patterns.

Fig. 1 c reproduces the DNA ridge pattern of Fig. 1 b and adds to it a set of solid lines lying in the valleys between the ridges. We picture these as the axes of a set of tubes, each made of many identical segments from different protein subunits, which are cradled by the grooves of the DNA. Alternately, one could begin with these coat-protein tubes and insert nucleotides into the chinks between them, again producing Fig. 1c.

The origin of the protein tubes is illustrated in Figs. 1d and e, in each of which a basic surface lattice for the helix of protein subunits, with pitch P_p, has been superimposed onto the tube axes. In these figures a protein subunit is represented by a dot, and the lattices of dots have been laid with rotational (R) and helical (H) symmetries in Fig. 1, d and e, respectively. Protein surface lattices of this sort, in which a subunit is represented as a straight line adjoining another subunit, have been given elsewhere (Banner et al., 1981; Makowski and Caspar, 1981). The central feature of this DNA-protein interaction is that each protein subunit contributes a segment of polypeptide that nestles into a valley defined by the DNA. Although belonging to different subunits, the successive polypeptide segments lying
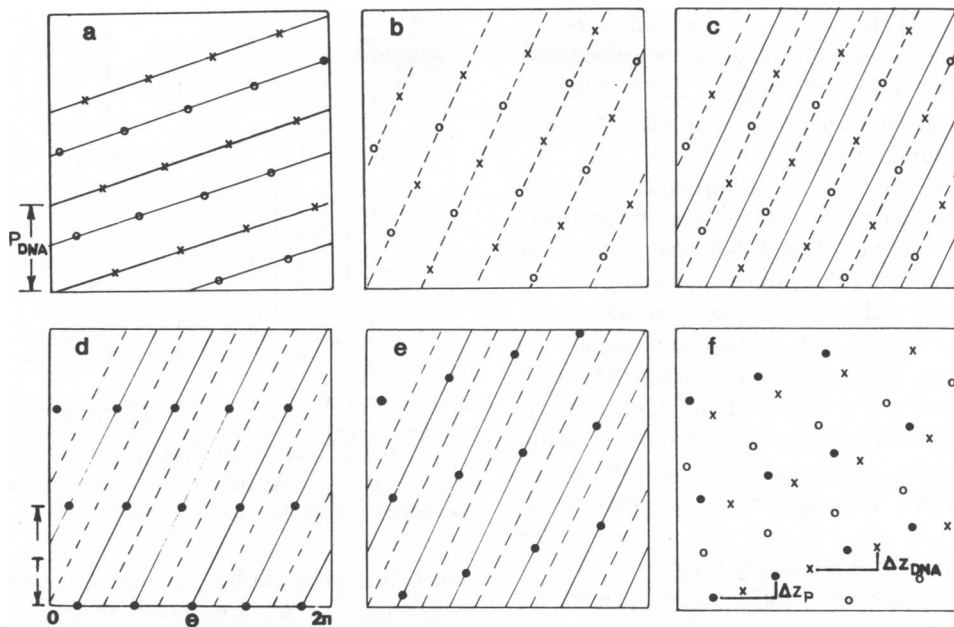
FIGURE 1   (a) An arbitrary helical DNA surface lattice, with equivalent points marked X (up strand) and 0 (down strand), and a solid line of pitch P_{DNA} schematically denoting the sugar phosphate backbone. (b) The backbone is removed. Dashed lines of pitch P join ridges of mass at large radius. (P equals twice the height of the figure.) (c) Segments of protein that form continuous helical tubes (solid lines) are inserted into the valleys between the ridge lines. (d) The nucleotides are removed and equivalent points on the protein subunits are denoted by dots, forming an example of a rotationally symmetrical surface lattice of subunits; N_R-mer axial translation T is shown. e, same as (d) but showing an example of a helically symmetrical protein surface lattice. f, the lattices of (e) Showing subunits and DNA but no ridge or valley lines; Δz_p and Δz_{DNA} are indicated.

along one of the tube axes, when taken together, form a continuous helical tube of protein that wraps around the virion along its entire length. A number, N, of such tubes of protein completely surround the DNA core. The remainder of each subunit extends into a second, outer layer of protein tubes coaxial with the original inner layer. Some relations between the two layers of protein, through which they form a double layer of close-packed tubes surrounding the DNA, will be presented by us elsewhere.

The geometry of Fig. 1 leads immediately to the pitch connection equation:

$$\frac{\Delta z_{DNA}}{P_{DNA}} = \frac{\Delta z_{DNA}}{P} + \frac{K}{N} \tag{1}$$

in which P_{DNA} is the pitch of the DNA helix, Δz_{DNA} is the nucleotide axial translation, and P is the pitch of a protein tube axis. K is a constant $= \pm1, \pm2, \pm1/2, \pm3, \pm1/3, \ldots$, the absolute value of which is the number of protein tubes per nucleotide, and the sign of which is $+(-)$ if successive nucleotides are associated with tubes having greater (smaller) azimuthal coordinate. Eq. 1 states that for a pitch connected DNA-protein interface, $\Delta\theta_{DNA}$ equals a contribution $(2\pi K)/N$ due to the azimuthal angle occupied by K of the N tubes plus a contribution $(\Delta z_{DNA})/P$ due to the axial rise per nucleotide along a helix tube of pitch P. Eq. 1 applies to helical DNA that is pitch connected to a protein coat that has either helical or rotational symmetry.

Note that the pitch P of the protein tube axes is distinctly different from the pitch P_p of the basic subunit lattice.

We denote the subunit axial translation by Δz_p, and so $n/s = 2\Delta z_p/\Delta z_{DNA}$. The factor of 2 arises because the filamentous bacteriophages locally have two strands of DNA. It is evident from Fig. 1, d and e, that the protein subunit axial translation Δz_p can be varied freely with the DNA lattice remaining fixed, simply by repositioning the dots along the tube axes. It is this freedom that allows a protein lattice with either H or R symmetry, and having any value of Δz_p, to be superimposed onto the tube axes of Fig. 1c. For the rotationally symmetrical case, $\Delta z_p = T/N_R$, where N_R is the order of rotational symmetry, and T is the axial rise between successive N_R-mers. Thus, any value of n/s can be accommodated by the pitch connection construction. Fig. 1 f shows the DNA and protein lattices of Fig. 1 e, without the ridge and valley lines.

THE RESTRICTED PITCH CONNECTION

In this section we consider the observed values of n/s as examples of pitch-connected symmetry. We begin with some general remarks and advance a hypothesis which affords selection of discrete sets of allowed n/s values that contain the observed examples.

We seek an optimum mode of DNA-protein interaction that is DNA-sequence independent. This is essential, since insertion and deletion mutations of the fd genome are

packaged indistinguishably from the native genome, except for a linear scaling of virion length with genome size (Enea and Zinder, 1975; Herrmann et al., 1978). Thus, in general, it is the mean shape of the DNA molecule that is fitted stereochemically to the protein coat.

It is plain that the local contribution to the total DNA-protein interaction varies on the DNA-protein interface, depending on what part of the DNA and what part of the protein are in contact. We assume that there is some one site on each nucleotide that interacts optimally with some corresponding site on each protein subunit, so that the net DNA-protein interaction is maximum when the largest number of DNA sites and protein sites are juxtaposed. The sites could be charged groups or aromatic residues for the protein, and phosphate charges or nucleotide bases for the DNA, although these examples are not exhaustive.

The simplest manner of achieving a large number of site-site interactions is through simple periodic symmetry, in which each nucleotide site (or each nth nucleotide site) binds to a subunit site, and every subunit site (or every nth subunit site) is bound to a nucleotide site. However, these symmetries do not allow for the observed nonintegral n/s ratios. To achieve both a large number of site-site interactions and a nonintegral n/s value, we invoke the pitch connection, to allow nonintegral n/s, and add to it the requirement that $n/s = 2j/k$, where j and k are integers. A rational value for n/s ensures that the number of different nucleotide environments is finite and that every jth nucleotide site and kth subunit site can be in proximity. A virion with N nucleotides would have N/j of them enjoying site-site interactions, so it is advantageous to keep j small; the same argument applies to k for the subunits. Fig. 2a shows an arbitrary example of restrictively pitch connected protein and DNA lattices, with $n/s = 2(7/5) = 2.80$. The dots here represent subunit sites, and the X's represent nucleotide sites, with each subunit site bracketed by two
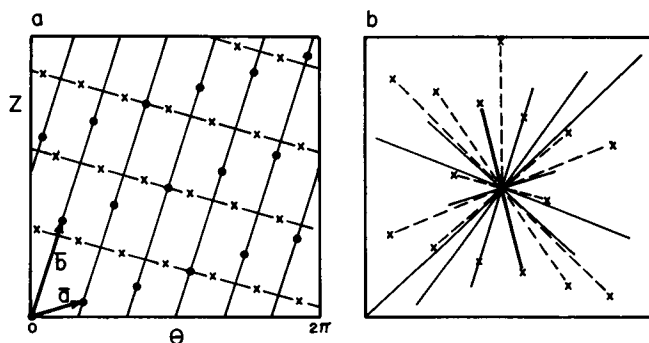

FIGURE 3 (a) A restrictively pitch-connected model of the DNA and protein surface lattice interactions of fd. $N_R = 5, M = 2, T = 16$ Å and $P_{DNA} = 26.7$ Å. (b) Restrictively pitch-connected surface lattices for Xf, with $N = 11$, $n/s = 2.0$, and every 11th subunit interacting strongly with two nucleotides. Both strands of the DNA are shown.

nucleotide sites. Other examples of restrictively pitch-connected lattices are shown in Figs. 3–5, drawn for application to the filamentous phages.

A rational value for n/s implies that the set of DNA lattice vectors and the set of protein lattice vectors have at least one common member. This is shown graphically in Fig. 2b, drawn for the lattices of Fig. 2a. The connection between rational n/s values and strongly interacting surface lattices is shown mathematically in the Appendix. The Appendix derives both the pitch connection equation and the restricted pitch connection by considering a generalized interaction between the surface lattice of a helical DNA molecule and the surface lattice of the protein coat. To derive the pitch connection, the protein coat is characterized as a set of N helices; to derive the restricted pitch connection, the discrete aspect of the protein surface lattice is emphasized. The restricted pitch connection results from a fusion of the two results: a rational n/s value and the pitch connection.

In drawing a pair of restrictively pitch-connected

FIGURE 2 (a) Restrictively pitch-connected DNA and protein lattices, with $n/s = 2(7/5)$ and $M = 1$. Every fifth subunit site interacts strongly with a pair of nucleotide sites. The primitive lattice vectors $\bar{a}$ and $\bar{b}$ for the protein helix are shown. (b) Solid and dashed lines denote lattice vectors of the protein DNA lattices, with the terminus of each DNA lattice vector marked by an X, and the heavy lines denoting common vectors.
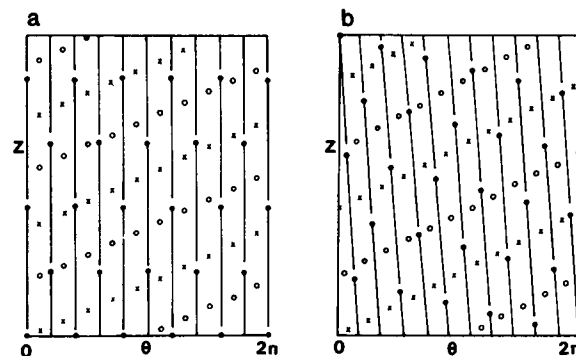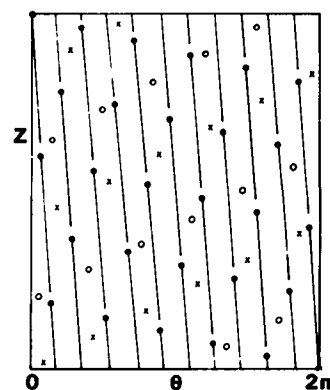

FIGURE 4 Restrictively pitch-connected surface lattices for Pf1, with $N = 11$, $n/s = 1.0$, and every 11th subunit interacting strongly with one nucleotide. The DNA ridges fit into every other protein groove in an alternate interaction mode. Both strands of DNA are shown.
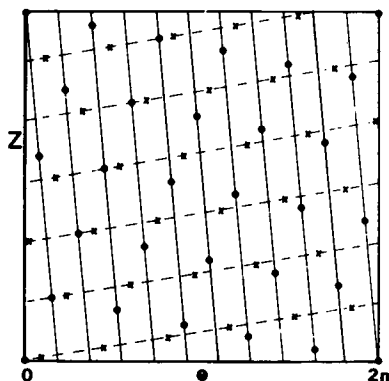
FIGURE 5 Restrictively pitch-connected surface lattices for Pf3, with $N = 11$, $n/s = 2.40$, and every 6th nucleotide interacting strongly with every 5th subunit. Both strands of DNA are shown.

lattices for a given number N of α-helix tubes the value of k in the expression $n/s = 2j/k$ can always be chosen arbitrarily. With N and k fixed, only a small set of values of j is possible; each j value corresponds to the number of tubes between one protein site and another that is k subunits further along the basic protein lattice. Hence the n/s spectrum is discrete, and its elements depend on the number of protein tubes surrounding the DNA. For an inclusive enumeration, these relations can be written algebraically.

For chosen values of N and k, the corresponding set of integers j is given by

$$j = \left| \frac{mN + k_0}{K} \right|. \tag{2}$$

The positive or negative index m can range over all integers, including zero; it is the number of complete turns (the sign of which is measured relative to the helical sense of the basic protein lattice) made by the DNA strand between successive site-site interactions minus the number of turns made by the basic protein lattice. Thus, if $m = 0$, the DNA helix follows that of the basic protein lattice. The integer k_0 is the number of tubes crossed by a DNA strand between successive site-site interactions. For helical lattices $k_0 = kM$, where the positive integer M is the number of cycles of the basic protein subunit helix cut by an α-helix tube axis as it spans the distance between its initial and terminal dots. Equivalently, M is the number of tubes between adjacent dots along the basic protein helix. In Fig. 1 d–e, $M = 1$, and in Figs. 3–5, $M = 2$.

In the case of a protein lattice possessing N_R-fold rotational symmetry, one finds $N = MN_R$. Because the protein N_R-mers occur at z intervals of $N_R \Delta z$, the allowed values of k are $k = LN_R$ for $L = 1,2, \ldots$. The allowed k_0 values are $k_0 = L + SM$, for $S = 0,1, \ldots, N_R - 1$. With these expressions in hand, a spectrum of n/s values can be predicted for a given protein surface lattice, once its subunits have been seated in a set of N helical tubes.

NMR studies by Cross et al. (1983) have found that the ^{31}P NMR spectrum for Pf1 shows chemical shifts for only one phosphate-group orientation, with respect to the structure axis, whereas the spectrum for fd shows many. These spectra are interpreted to mean that the sugar-phosphate dihedral angles of Pf1 are roughly constant from nucleotide to nucleotide, whereas those of fd vary. However, a laser Raman spectroscopy study of fd (Thomas et al., 1983) finds that the sugar-phosphate backbone of fd is well ordered. The restricted pitch connection explains how the fd DNA might appear ordered by some techniques (x-ray, Raman spectroscopy, and IR), yet disordered by another (NMR). By connecting every jth nucleotide to every kth subunit, the restricted pitch connection actually relates the protein coat only to the average structure of the DNA, where the average is taken over j successive nucleotides. Thus the $j - 1$ nucleotides lying between successive nucleotide sites can assume local helix parameters, varying from the average values, to adjust to the local protein configuration. Perturbations in Δz_{DNA} merely cause the nucleotides to slide up or down along their protein grooves, in accord with the unrestricted pitch connection hypothesis. It is possible for the sequence of DNA-protein interactions, and therefore the values of the dihedral angles of the sugar-phosphate backbone, to repeat exactly every j nucleotides; then the nucleotide-protein interaction would have a unit cell containing j nucleotides and k subunits. The x-ray photos reflect the helix parameters of the average DNA structure, whereas the NMR spectra are sensitive to the j different local nucleotide configurations, at the level of dihedral angles.

APPLICATION TO fd, Xf, Pf1, AND Pf3

We now apply the restricted pitch connection to four of the filamentous viruses, beginning with fd. Its fivefold rotational symmetry and 16-Å pentamer translation are taken from x-ray diffraction studies (Makowski and Caspar, 1978; Banner et al., 1981; Makowski and Caspar, 1981), and so $\Delta z_p = (16/5)$ Å $= 3.2$ Å. X-ray photos of magnetically oriented fd fibers show no layer line splitting and contain two spots, assigned to the DNA, at ~1/26.7 Å^{-1} and 2/26.7 Å^{-1} (Banner et al., 1981). The absence of layer line splitting indicates an exact 36° rotation angle between pentamers, as shown in the dot lattice of Fig. 3 a. The fd DNA lattice of this figure is drawn with right-handed pitch, in accord with the arguments presented by Casadevall and Day (1982). The DNA spots indicate that $P_{DNA} = 26.7$ Å in fd, an experimental fact to be interpreted.

To do this we apply Eq. 1 and the rotational case of Eq. 2:

$$j = \left| \frac{mMN_R + L + SM}{K} \right|. \tag{3}$$

We seek indices m, L, S, M, K, and k (with $k = 5L$) such that n/s falls in the range 2.3–2.4. We also need P_{DNA}, found from Eq. 1 and $\Delta z_{DNA} = k\Delta z_p/j$, to be near 26.7 Å. To apply Eq. 1 we need P. Our modeling experiments indicate that the only possible values for P consistent with the x-ray data are $P = 160$ Å for $M = 1$ and $P = \infty$ for $M = 2$. A systematic search reveals that it is possible to find $j/k = 2.40$ in two ways, but that only one solution yields an appropriate P_{DNA}. This solution has $j = 12$, $k = 10$, $m = 2$, $K = 1$, $P = \infty$, and $P_{DNA} = +10(3.2/1.2) = +26.67$ Å. The nearest rejected P_{DNA} values are 23 and 32 Å; the nearest rejected n/s values are 2.2 and 2.6. These results could have been obtained graphically, but with considerably more effort.

The acceptable solution has the DNA with 10 residues per turn, surrounded by 10 exactly vertical tubes of protein. This structure is in accord with linear dichroism data (Fritzche et al., 1981; T. Troxel and J. A. Schellman, personal communication), NMR data (Cross et al., 1983), x-ray data (Marvin et al., 1974b), and a magnetic birefringence study (Torbet and Maret, 1981), which indicate that the α-helices of fd are nearly parallel to the structure axis. In a subunit packing study to be presented elsewhere, we have developed this model of fd and used it to calculate diffraction patterns in agreement with the semiquantitative patterns available in the literature (Marvin et al., 1974b). We propose in view of this evidence, that $n/s = 2.40$ exactly for fd.

We turn now to Xf and Pf1. Their very similar x-ray diffraction patterns (Marvin et al., 1974a) indicate that their protein coats are also similar on a scale of ~ 10 Å and that both have helical symmetry. Their DNA morphologies differ significantly, since Xf has an experimental n/s value nearly equal to 2, whereas Pf1 has an n/s value nearly equal to 1 (Wiseman and Day, 1977; Day and Wiseman, 1978). We shall assume that the correct values are 2 and 1, respectively, and consider the consequences attached to explaining these numbers via the simple periodic symmetry assumption and via the restricted pitch connection.

With integral n/s values, a pitch-connected DNA-protein interface is possible but not required, since an interaction with simple periodic symmetry is sufficient, and in fact, has been accepted as a matter of course in the past. However, a simple periodic DNA-protein symmetry implies that the DNA lattice and basic protein lattice share the same pitch, known from x-ray studies to be 15 Å. There is no intrinsic problem with such a low pitch for the DNA of Xf and Pf1 (Day et al., 1979). However, silver binding studies (Casadevall and Day, 1982) strongly suggest that the DNAs of native Xf and fd are very similar. Because fd DNA in its virion has 27-Å pitch, a contradiction is implied by the assumption of simple periodic symmetry: Xf DNA in its virion has both a 15-Å pitch and an ~ 27-Å pitch. We can use the restricted pitch connection hypothesis to dissolve this contradiction by voiding the simple periodic

symmetry assumption for Xf and its concommitant 15-Å DNA pitch.

To see how integer n/s values proceed from Eq. 2, we apply it to the case of helical protein symmetry:

$$j = \left| \frac{mN + kM}{K} \right|. \tag{4}$$

The sign of $mN + kN$ indicates whether the DNA and the protein tube helices have the same $(+)$ or opposite $(-)$ handedness. If $m = 0$, the solution has simple periodic symmetry. Solutions with $n/s = 2.0/K$ can be found by setting $j = k = N$ and $m + M = \pm 1$. Every value of N thus possesses these two solutions with $n/s = 2.0/K$, so these values of n/s can be thought of as being generic. The two solutions with $n/s = 2.0$ and $K = 1$ are plainly related to solutions with $n/s = 2.0/K$ for $K \neq 1$.

To apply Eq. 4 to Xf and Pf1, we need values of M and N. In our theoretical study of protein packing in filamentous phages, we have found two Pf1 (and Xf) models that agree with the experimental fiber diffraction patterns, and these have $(M,N) = (2,11)$ and $(M,N) = (1,6)$. From the pitch connection Eq. 1, together with $n/s = 2.0$, $\Delta z_p = 2.77$ Å for Xf, and several K values, we have calculated P_{DNA} for these models. The value of P needed for Eq. 1 is fixed by the lattice parameters of Xf (Table I) and the value of N. The Pf1 and Xf rotation angles $\Delta\theta_p$ have absolute magnitudes of $(5/27)2\pi$. For Pf1 the sign of $\Delta\theta_p$ is thought to be negative (Makowski et al., 1980), and because of the similarity of the Pf1 and Xf diffraction patterns we assume this is the case for Xf as well. Taking $N = 11$, we find for the pitch of the alpha helices in the protein layer surrounding the DNA, $P = 2\pi[(11\Delta z_p)/(11\Delta\theta_p + 4\pi)] = -825$ Å. Setting $K = 1$, we find for the $(M,N) = (2,11)$ models $P_{DNA} = 31.6$ Å, with $j = k = 11$ and $m = -3$. The (1,6) model requires $P = -150$ Å, and the positive DNA pitches given by the pitch connection equation are, for $K = 1$, $P_{DNA} = 18.7$ Å, and, for $K = 1/2$, $P_{DNA} = 42.8$ Å. Because these are the only positive DNA pitches, the six-tube solution is rejected by the pitch connection equation, whereas the 11-tube solution allows a DNA pitch close to the 27-Å pitch observed for fd. For this reason and because of its similarity to the fd model, about which there is little or no ambiguity, we favor the $N = 11$ solution for Xf and Pf1. Another line of reasoning discredits the $N = 6$ models. These solutions have protein tubes with a relatively small pitch, 150 Å in absolute value, and correspondingly large curvature, so they may be energetically unfavorable, since α-helices are fairly rigid. The standard x-ray fiber-diffraction pattern analysis of Makowski et al. (1980) also indicates for Pf1 a $(2,11)$ type structure, although they do not use this notation. Fig. 3b shows the protein and DNA surface lattices, restrictively pitch connected for the Xf model with $N = 11$, $n/s = 2.0$, $K = 1$, $m = -3$, and $j = k = 11$.

For Pf1 we expect the same large-scale protein configu-

ration as in Xf but exactly one-half of the nucleotides, since it has $n/s = 1.0$ and $\Delta z_{DNA} = 5.5$ Å. To arrange this we can adapt an Xf DNA-protein lattice by simply removing every other nucleotide. The resulting lattice is shown in Fig. 4. The "alternate interaction" character of this DNA-protein interface, wherein every other groove between tubes is skipped, is handled algebraically by setting $K = 2$. The alternate interaction mode is not an arbitrary device. Because there are two strands of DNA, even the simplest possible Pf1 DNA-protein interface, with a simple periodic interaction, would necessarily have successive nucleotides of one strand associated with every other subunit along the basic protein lattice. Although the DNA lattice of Fig. 4 formally has 22 different protein environments for its nucleotides, from the standpoint of the proximity required for site-site interactions, there are only 11, just as for Xf.

We have shown elsewhere (Day et al., 1979) that the large Δz_{DNA} of Pf1 mandates an inverted DNA configuration (I-DNA) in which the bases are directed outward, away from the structure axis, and the phosphates are located near the structure axis. This result holds for Pf1 DNA models with 15-Å or ~30-Å pitches. However, a stereochemical feasibility computation with the linked-atom least-squares computer code demonstrates the intuitively acceptable fact that the 30-Å pitch model is less tightly wound than the 15-Å pitch model. The 30-Å pitch models have much smaller electron density on the helix axis, and most atoms are at larger radii, allowing a more relaxed structure. Because the sugar-phosphate backbone of an I-DNA structure is more isolated from the protein than that of a classical DNA structure, it could escape structural perturbations induced by a varying protein environment and have constant dihedral angles. This appears to be the case in the NMR spectra observed for Pf1.

The Pf3 x-ray diffraction pattern has not yet been thoroughly analyzed. It is known that its protein coat has helical symmetry and the same lattice parameters Δz_p and $\Delta \theta_p$ as Pf1 and Xf (Peterson et al., 1982). The n/s ratio of Pf3 is 2.38 ± 0.14 (Table I), so we now apply Eq.4, seeking j and k values that give n/s within the above limits. In the range $k = 1$ to $k = 16$ we find several possible solutions, with $N = 5$ to 14. Of these, the most plausible has $N = 11$, for the same reasons as apply to Xf: it allows rigid α-helices and good comparison with fd. This solution has $K = 2$, $M = 2$, $j = 6$ and $k = 5$, yielding $n/s = 2.40$ and, from Eq. 1, $P_{DNA} = 12.9$ Å. It is shown in Fig. 5. This low P_{DNA} value requires an I-DNA structure for Pf3, but because of its small Δz_{DNA} value, the Pf3 I-DNA need not be as tightly wound as a Pf1 DNA model with 15-Å pitch.

Alternatively, if the range of Pf3 n/s values were unknown, setting $N = 11$ would severely constrain the spectrum of n/s values. These may be calculated from Eq. 4 by inserting all possible values of k, m, and K using $N = 11$ and $M = 2$. The $K = 1/2$ choice produces small $\Delta \theta_{DNA}$ values of ~18° and, consequently, crowded nucleotides, so

TABLE II
POSSIBLE RESTRICTIVELY CONNECTED LATTICES WITH ELEVEN TUBES FOR Pf3

K	± 1	± 1	± 2	± 2	± 2
m	-1	-2	-2	-2	-2
j	3	8	6	5	4
k	4	7	5	6	7
n/s	1.5	2.28	2.40	1.67	1.14
P_{DNA}	42.9 Å	27.6 Å	12.9 Å	18.8 Å	27.6 Å
	-38.8 Å	-25.9 Å	-12.5 Å	-17.9 Å	-25.9 Å

Solutions to Eq. 4 with $(M,N) = (2,11)$, $j < 9$, and $K = \pm 1$ and ± 2, which yield $0.5 < n/s < 3.0$. P_{DNA} is calculated from Eq. 1, with $\Delta z_{DNA} = k\Delta z_p/j$ and $P = -825$ Å.

the $K = 1/2$ solutions can be discarded. All remaining solutions with $0.5 < n/s < 3.0$ are listed in Table II. The $m = 0$ solution, for $j = k = $ any integer, is a simple periodic symmetry solution, as remarked above. Table II contains only one entry with j and k values smaller than those we propose for Pf3, and its n/s value is quite removed from 2.40. In this way, the 11-tube structural constraint might explain why the value 2.40 appears for Pf3, instead of neighboring numbers.

For each of Xf, Pf1, and Pf3, $(M,N) = (1,6)$ and $(M,N) = (2,11)$ models are feasible on the basis of subunit packing constraints, diffraction pattern analyses and agreement of experimental n/s values with those allowed by the restricted pitch connection. However, in each case the $(M,N) = (2,11)$ models are arguably favorable.

Comparison of Figs. 3 b, 4, and 5 reveals strong similarities in the possible DNA-protein interactions they represent. The choice $N = 11$ is built into the three protein configurations, but it is a striking fact that for each of Xf, Pf1, and Pf3, 11-tube models are possible that satisfy Eq. 4 for successful n/s values. It is also noteworthy that the Pf1 and Pf3 models both require an I-DNA configuration and an alternate interaction $(K = 2)$ between DNA and protein, perhaps suggesting a link between these features. If the Pf3 and fd models are correct, it may be of considerable significance that both have $n/s = 2.40$ exactly. For these reasons and their common mass-per-length values (Table I), the fd, Xf, Pf1, and Pf3 structures can plausibly be considered as variations on one successful theme, and this view might carry implications for the study of their evolutionary history.

CONCLUSION

The pitch connection and restricted pitch connection presented above were developed to resolve the theoretical problems posed by the nonintegral n/s values of fd and Pf3. In addition they resolved the contradictions implied by the silver binding study of Xf and suggested a quantitative theoretical basis for the fd DNA reflections observed in x-ray diffraction patterns. The DNA pitches of Xf and Pf1 are predicted, with an alternative for Pf1 to the tightly

wound 15-Å pitch DNA demanded by the straightforward simple periodic interaction hypothesis. Finally, the restrictive pitch connection hypothesis predicts that a mutation that changes the n/s value of a virus must change it by a discrete jump, with a wholesale change in the symmetry and stability of the DNA-protein interface.

As a means of rectifying evidently disparate symmetries, the pitch connection and restricted pitch connection may describe the relations between coaxial helices in other biological structures.

APPENDIX

The pitch connection and the restricted pitch connection will be derived here more formally by considering the mutual interactions of protein and DNA lattices. To represent the lattices we need some notation. With primitive lattice vectors $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ as shown in Fig. 2 a, we define reciprocal space vectors $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ by

$$\tilde{\mathbf{A}} \cdot \bar{\mathbf{a}} = \tilde{\mathbf{B}} \cdot \bar{\mathbf{b}} = 0 \tag{A1a}$$

and

$$\tilde{\mathbf{A}} \cdot \bar{\mathbf{b}} = \tilde{\mathbf{B}} \cdot \bar{\mathbf{a}} = 2\pi . \tag{A1b}$$

Conditions A1 lead to

$$\tilde{\mathbf{A}} = \frac{2\pi}{a^2 b^2 - (\bar{\mathbf{a}} \cdot \bar{\mathbf{b}})^2} [-(\bar{\mathbf{a}} \cdot \bar{\mathbf{b}})\bar{\mathbf{a}} + a^2 \bar{\mathbf{b}}] \tag{A2a}$$

and

$$\tilde{\mathbf{B}} = \frac{2\pi}{a^2 b^2 - (\bar{\mathbf{a}} \cdot \bar{\mathbf{b}})^2} [-(\bar{\mathbf{a}} \cdot \bar{\mathbf{b}})\bar{\mathbf{b}} + b^2 \bar{\mathbf{a}}] . \tag{A2b}$$

For a lattice wrapped around a cylinder, the vectors $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ are related by the periodicity condition

$$\bar{\mathbf{b}} = \bar{\mathbf{e}}_x [La_x - 2\pi R \operatorname{sgn}(a_x)] + \bar{\mathbf{e}}_z(La_z) . \tag{A3}$$

Here L is the integer part of $[(2\pi R/a_x) + |a_x|/a_x]$, and $\bar{\mathbf{e}}_x$ and $\bar{\mathbf{e}}_z$ are unit vectors in the azimuthal and structure axis directions, respectively. Now, a physically realizable function f of position $\bar{\mathbf{s}} = x\bar{\mathbf{e}}_x + z\bar{\mathbf{e}}_z$ on the lattice may be written as a Fourier series

$$f(\bar{\mathbf{s}}) = \sum_{p,q=-\infty}^{\infty} C_{pq} e^{i\tilde{\mathbf{k}}_{pq} \cdot \bar{\mathbf{s}}} \tag{A4}$$

with $\tilde{\mathbf{k}}_{pq} = p\tilde{\mathbf{A}} + q\tilde{\mathbf{B}}$ and

$$C_{pq} = \frac{1}{2\pi RH} \int_{x=0}^{2\pi R} dx \int_{z=0}^{H} dz \, e^{-i\tilde{\mathbf{k}}_{pq} \cdot \bar{\mathbf{s}}} f(\bar{\mathbf{s}}), \tag{A5}$$

where H is the length of the virion. Eqs. A4 and A5 ensure periodicity in the representation of the lattice

$$f(\bar{\mathbf{s}}_{lm} + \bar{\mathbf{e}}) = f(\bar{\mathbf{e}}),$$

where $\bar{\mathbf{s}}_{lm} = l\bar{\mathbf{a}} + m\bar{\mathbf{b}}$. Eqs. A1 and A3 lead to

$$(k_{pq})_x = \frac{(Lq - p) \operatorname{sgn}(a_x)}{R} , \tag{A6}$$

which ensures periodicity around the cylinder

$$f(\bar{\mathbf{s}} + 2\pi R\bar{\mathbf{e}}_x) = f(\bar{\mathbf{s}}).$$

To derive the pitch connection we shall write the interaction between a helix of nucleotides having pitch P_{DNA} and N continuous helices of protein having pitch P, coaxial with the DNA. We represent the DNA by a function $F^D(\bar{\mathbf{s}})$, as in Eq. A4, with Fourier coefficients C_{pq}. This quantity is meant to denote some property defined on the DNA surface lattice through which the DNA interacts with the protein. We define F^D in order to write the area density of the interaction as the product of $F^D(\bar{\mathbf{s}})$ with a similar quantity $F^P(\bar{\mathbf{s}})$, defined for the protein. For example, we might take F^D and F^P to be the DNA charge density and the protein electrostatic potential, respectively.

The protein surface function is written

$$F^P(\bar{\mathbf{s}}) = \sum_{j=-\infty}^{\infty} D_j e^{ijN[(x/R)-(2\pi z/P)]}. \tag{A7}$$

This form ensures that the representation of the protein has N-fold helical symmetry by meeting the conditions

$$F^P\left(\bar{\mathbf{s}} + \frac{2\pi R\bar{\mathbf{e}}_x}{N}\right) = F^P(\bar{\mathbf{s}})$$

and

$$F^P \left| \bar{\mathbf{s}} + \left(\frac{2\pi R}{P}\bar{\mathbf{e}}_x + \bar{\mathbf{e}}_z\right)z \right| = F^P(\bar{\mathbf{s}}).$$

The DNA Fourier coefficients C_{pq} have two indices, whereas the protein tube Fourier coefficients D_j have only one. This is due to the fact that the value of $F^D(\bar{\mathbf{s}})$ depends on the position of $\bar{\mathbf{s}}$ with respect to a net of lattice points (representing the nucleotides), while the value of $F^P(\bar{\mathbf{s}})$ depends on the position of $\bar{\mathbf{s}}$ with respect to a set of lines (representing the protein tube axes). The protein lattice–DNA lattice interaction is written as

$$I = \int_0^{2\pi R} dx \int_0^H dz \, F^P(\bar{\mathbf{s}}) F^D(\bar{\mathbf{s}}). \tag{A8}$$

Substituting Eqs. A4 [with $F^D = f(\bar{\mathbf{s}})$] and A7 into Eq. A8 and integrating reveals that the term $C_{pq}D_j$ makes a nonvanishing contribution to I only if

$$\frac{jN}{R} + (k_{pq})_x = 0 \tag{A9a}$$

and

$$\frac{-2\pi jN}{P} + (k_{pq})_z = 0. \tag{A9b}$$

From Eqs. A2, A3, and the definition of $\tilde{\mathbf{k}}_{pq}$ we find

$$(k_{pq})_z = \frac{2\pi q}{a_z} + \frac{(p - qL)|a_x|}{Ra_z} . \tag{A10}$$

Substituting Eq. A10 into A9b and Eq. A6 into A9a and solving for $(Lq - p)$ yields

$$\frac{a_x}{2\pi R} + \frac{q/j}{N} = \frac{a_z}{P} . \tag{A11}$$

Because $a_z = \Delta z_{DNA}$ and $P_{DNA} = (2\pi Ra_z)/a_x$, we have

$$\frac{\Delta z_{DNA}}{P_{DNA}} + \frac{K}{N} = \frac{\Delta z_{DNA}}{P} , \tag{A12}$$

where $K = q/j$. This is the pitch connection equation; it results from a spatial resonance between one term in the Fourier expansion of the DNA surface function F^D and one term in the Fourier expansion of the protein tube surface function F^P.

The restricted pitch connection can be derived similarly. For this purpose both the DNA and the protein surface lattice functions are represented by Eq. A4, with Fourier coefficients C_{pq} and E_{lm}, respectively. Here the protein is treated as a lattice of points, each of which represents a site on a subunit that we suppose interacts strongly with some corresponding point on a nucleotide. Although the DNA must be helical, the lattice of protein points may have either helical or rotational symmetry. We write the interaction integral I as in Eq. A8 and note that the term $C_{pq} E_{lm}$ contributes only if

$$\tilde{\mathbf{k}}^D_{pq} = -\tilde{\mathbf{k}}^P_{lm} = \tilde{\mathbf{k}}^P_{-l,-m}. \tag{A13}$$

Eq. A13 states that the set of DNA reciprocal space vectors $|\tilde{\mathbf{k}}^D|$ and the set of protein reciprocal space vectors $|\tilde{\mathbf{k}}^P|$ have an element in common. It is straightforward to show that $\tilde{\mathbf{k}}^D_{pq} = \tilde{\mathbf{k}}^P_{-l,-m}$ implies $\bar{s}^D_{p,q} = \bar{s}^P_{l,m}$; i.e., the protein and DNA surface lattices share a common lattice vector. This is equivalent to the assertion that lattice points $\bar{s}^D_{-jp,jq}$ and $\bar{s}^P_{jl,-jm}$ are in strong interaction position, for all integers j. When this condition and the pitch connection equation hold simultaneously, we say the restricted pitch connection holds.

Eqs. A9 have the trivial solution $p = q = j = 0$; likewise, Eq. A13 has a solution $p = q = l = m = 0$. If these are the only solutions to Eqs. A9 and A13, then the DNA and protein interact in a completely irregular manner, relating to each other only through their mean properties, with surviving interaction terms $C_{00}D_0$ and $C_{00}E_{00}$, respectively. These terms are also present, of course, in a (restricted) pitch-connected structure, so the mean interaction is not lost when the interaction mediated by the (restricted) pitch connection is gained.

REFERENCES

Banner, D., C. Nave, and D. A. Marvin. 1981. A structure of protein and DNA in fd filamentous bacterial virus. *Nature (Lond.).* 289:814–816.

Berkowitz, S. A., and L. A. Day. 1980. Turbidity measurements in an analytical ultracentrifuge. Determinations of mass-per-length for filamentous viruses fd, Xf, and Pf3. *Biochemistry.* 19:2696–2702.

Casadevall, A., and L. A. Day. 1982. DNA packing in the filamentous viruses fd, Xf, Pf1 and Pf3. *Nucleic Acids Res.* 10:2467–2481.

Chothia, C., M. Levitt, and D. Richardson. 1977. Packing of α-helices and pleated sheets. *Proc. Natl. Acad. Sci. USA.* 74:4130–4134.

Cross, T. A., P. Tsang, and S. J. Opella. 1983. Comparison of protein and DNA backbone structures in fd and Pf1 bacteriophages. *Biochemistry.* 22:721–726.

Day, L. A., and R. L. Wiseman. 1978. A comparison of DNA packaging in the virions of fd, Xf, and Pf1. *In* The Single-stranded DNA Phages. D. Denhart, D. Dressler, and D. S. Ray, editors. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 605–625.

Day, L. A., R. L. Wiseman, C. J. Marzec. 1979. Structure models for

DNA in filamentous viruses with phosphates near the center. *Nucleic Acids Res.* 7:1393–1403.

Enea, V., and N. D. Zinder. 1975. A deletion mutant of bacteriophage f1 containing no intact cistrons. *Virology.* 68:105–114.

Fritzsche, H., T. A. Cross, S. J. Opella, and N. R. Kallenbach. 1981. Structure and architecture of the bacterial virus fd. An infrared linear dichroism study. *Biophys. Chem.* 14:283–291.

Grant, R. A., T.-C. Lin, W. Konigsberg, and R. E. Webster. 1981. Structure of filamentous bacteriophage f1. Location of the A, C, and D minor coat proteins. *J. Biol. Chem.* 256:539–546.

Herrmann, R., K. Neugebauer, H. Zentgraf, and H. Schaller. 1978. Transposition of a DNA sequence determining kanamycin resistance into the single-stranded genome of bacteriophage fd. *Mol. Gen. Genet.* 159:171–178.

Makowski, L., and D. L. D. Caspar. 1978. Filamentous bacteriophage Pf1 has 27 subunits in its axial repeat. *In* The Single-stranded DNA Phages. D. Denhardt, D. Dressler, and D. S. Ray, editors. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 627–643.

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

Makowski, L., D. L. D. Caspar, and D. A. Marvin. 1980. Filamentous bacteriophage Pf1 structure determined at 7-A resolution by refinement of models for the α-helical subunit. *J. Mol. Biol.* 140:149–181.

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

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

Marzec, C. J., and L. A. Day. 1980. DNA and protein packing in type I filamentous bacteriophage. *Biophys. J.* 32:240–242.

Nave, C., R. S. Brown, A. G. Fowler, J. E. Ladner, and D. A. Marvin. 1981. Pf1 filamentous bacterial virus x-ray fibre diffraction analysis of two heavy-atom derivatives. *J. Mol. Biol.* 149:675–707.

Newman, J., H. L. Swinney, and L. A. Day. 1977. Hydrodynamic properties and structure of fd virus. *J. Mol. Biol.* 116:593–606.

Newman, J., L. A. Day, G. Dalack, and D. Eden. 1982. Hydrodynamic determination of the molecular weight, dimensions, and structural parameters of Pf3 virus. *Biochemistry.* 21:3352–3358.

Peterson, C., G. W. Dalack, L. A. Day, and W. T. Winter. 1982. Structure of the filamentous bacteriophage, Pf3, by x-ray fiber diffraction. *J. Mol. Biol.* 162:877–881.

Thomas, G. J., Jr., and L. A. Day. 1981. Conformational transitions in Pf3 and their implications for the structure and assembly of filamentous bacterial viruses. *Proc. Natl. Acad. Sci.* 78:2962–2966.

Thomas, G. J., Jr., B. Prescott, and L. A. Day. 1983. Structure similarity, difference and variability in the filamentous viruses fd, If1, IKe, Pf1, Xf and Pf3: investigation by laser Raman spectroscopy. *J. Mol. Biol.* In press.

Torbet, J., and G. Maret. 1981. High field magnetic birefringence study of the structure of rodlike phages in solution. *Biopolymers.* 20:2657–2669.

Webster, R. E., R. A. Grant, and L. A. W. Hamilton. 1981. Orientation of the DNA in the filamentous bacteriophage f1. *J. Mol. Biol.* 152:357–374.

Wiseman, R. L., and L. A. Day. 1977. Different packaging of DNA in the filamentous viruses Pf1 and Xf. *J. Mol. Biol.* 116:607–611.