

Available online at www.sciencedirect.com





### Structure of the Coat Protein in Pf1 Bacteriophage Determined by Solid-state NMR Spectroscopy

# David S. Thiriot, Alexander A. Nevzorov, Lena Zagyanskiy, Chin H. Wu and Stanley J. Opella\*

Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0307, USA

The atomic resolution structure of Pf1 coat protein determined by solidstate NMR spectroscopy of magnetically aligned filamentous bacteriophage particles in solution is compared to the structures previously determined by X-ray fiber and neutron diffraction, the structure of its membrane-bound form, and the structure of fd coat protein. These structural comparisons provide insights into several biological properties, differences between class I and class II filamentous bacteriophages, and the assembly process. The six N-terminal amino acid residues adopt an unusual "double hook" conformation on the outside of the bacteriophage particle. The solid-state NMR results indicate that at 30 °C, some of the coat protein subunits assume a single, fully structured conformation, and some have a few mobile residues that provide a break between two helical segments, in agreement with structural models from X-ray fiber and neutron diffraction, respectively. The atomic resolution structure determined by solid-state NMR for residues 7-14 and 18-46, which excludes the N-terminal double hook and the break between the helical segments, but encompasses more than 80% of the backbone including the distinct kink at residue 29, agrees with that determined by X-ray fiber diffraction with an RMSD value of 2.0 Å. The symmetry and distance constraints determined by X-ray fiber and neutron diffraction enable the construction of an accurate model of the bacteriophage particle from the coordinates of the coat protein monomers.

© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* Pf1; filamentous bacteriophage; protein structure; solid-state NMR

\*Corresponding author

#### Introduction

Filamentous bacteriophage particles consist of several thousand symmetrically arranged coat protein subunits that form a protective tube around the single-stranded circular DNA which is stretched lengthwise.  $^{1.2}$  Structural studies of filamentous bacteriophages remain outside the realm of conventional experimental methods: Pf1 bacteriophage particles have not been crystallized for X-ray diffraction, and their mass (>30,000 kDa) and asymmetric shape (60 Å  $\times$  20,000 Å) preclude the rapid isotropic reorientation essential for solu-

Abbreviations used: RMSD, root-mean-square deviation; ppm, parts per million; PDB, Protein Data Bank; PISA, polarity index slant angle.

E-mail address of the corresponding author: sopella@ucsd.edu

tion NMR spectroscopy. As a result, Pf1 has played important roles in the development of methods for structure determination of biosuited logical supramolecular assemblies. X-ray fiber diffraction,<sup>3–6</sup> neutron diffraction,<sup>7,8</sup> laser Raman spectroscopy,<sup>9,10</sup> and our earlier solid-state NMR studies<sup>11,12</sup> have contributed complementary (albeit to some degree conflicting) structural information about the major coat protein at moderate to high resolutions. <sup>31</sup>P NMR spectra of aligned bacteriophage particles show that the DNA is completely ordered with each phosphodiester linkage having the same orientation with respect to the long axis of the filament and the magnetic field, as expected for a 1:1 ratio of nucleotides to coat protein subunits.11 Pf1 coat protein also has a second significant biological role as an integral membrane protein. Comparisons of the structure and dynamics of Pf1 coat protein in membrane

environments $^{12-17}$  and in bacteriophage particles provide insights into the structural rearrangement that accompanies viral assembly. $^{15,18,19}$ 

As demonstrated with fd (PDB entry 1NH4),20 the backbone structure of the coat protein in a filamentous bacteriophage particle can be determined from the orientational constraints provided by the <sup>15</sup>N chemical shift and <sup>1</sup>H-<sup>15</sup>N dipole-dipole coupling frequencies measured from two-dimensional solid-state NMR spectra. In this context, the term "solid-state" refers to the experimental methods and instrumentation used to obtain highresolution NMR spectra of proteins that do not undergo large-amplitude motions on timescales that are short  $(10^{-4} \text{ Hz})$  compared to the inverse of the spectral range from the operative chemical shift and heteronuclear dipole-dipole coupling interactions. The samples themselves are concentrated solutions of virus particles. Structural fitting<sup>21</sup> of assigned experimental NMR data yields a unique three-dimensional structure of the protein. For the Pf1 coat protein structure described here (PDB entry 1PJF), the effective resolution determined by taking into account all sources of experimental error and uncertainties in the spin interaction tensors is estimated to be 1.5 Å, which compares favorably to the values reported for the structures determined by X-ray<sup>5</sup> fiber and neutron<sup>7,8</sup> diffraction. At atomic resolution, the structure of Pf1 coat protein is remarkably complex. In addition to the striking double hook of the N-terminal region, the predominantly  $\alpha$ -helical protein has a mobile break centered at residue 16, as suggested by neutron diffraction,<sup>7,8</sup> and a distinct kink at residue 29, which has also been found in a highly refined X-ray fiber diffraction structure (PDB entry 1QL1).5

#### Results

#### NMR spectroscopy of Pf1 virus particles

The two-dimensional NMR spectrum of uniformly 15N-labeled Pf1 bacteriophage shown in Figure 1(a) contains resolved resonances from all backbone and side-chain nitrogen sites of the coat protein. These resonances are spread over the full range of <sup>1</sup>H and <sup>15</sup>N chemical shift frequencies and the data were obtained at a relatively high <sup>1</sup>H resonance frequency (750 MHz); therefore, in order to have all resonances represented in a single spectrum it was necessary to make a composite from data obtained with two different high resolution separated local field experiments. 17,18 All spectral features apparent in Figure 1(a) have been confirmed in spectra of uniformly 15N-labeled samples obtained under slightly different experimental conditions, e.g. offset frequencies or radio frequency power levels, or in those of selectively <sup>15</sup>N-labeled samples, as shown in Figure 1(d).

The resonance assignments were obtained with a modified version of the shotgun NMR approach

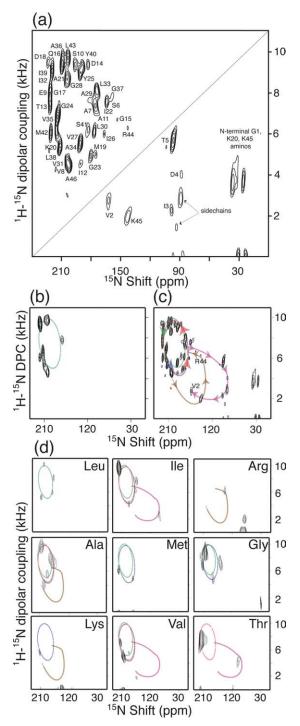
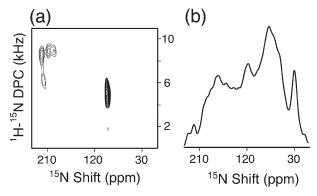


Figure 1. Two-dimensional separated local field spectra of Pf1 bacteriophage particles. (a) Composite spectrum of uniformly 15N-labeled Pf1 derived from a PISEMA spectrum (upper left triangle) and a SAMMY spectrum (lower right triangle). Resonance assignments are indicated. (b)-(d) Color coding by residue shows simulated PISA wheels or arcs for the helical and N or C-terminal regions, respectively, according to the following scheme: magenta, Gly1-Ser6; red, Ala7-Asp14; blue, Gly17-Gly 28; green, Leu30-Leu43; brown, Arg44-Ala46. (b) PISEMA spectra of a uniformly 15N-labeled sample after two days of <sup>2</sup>H<sub>2</sub>O exchange. (c) Same spectrum as in (a) marked with colored lines and arrows to illustrate the PISA wheels and sense of rotation of the resonances, counting back from C terminus to N terminus. (d) PISEMA spectra of selectively <sup>15</sup>N-labeled samples.



**Figure 2.** Solid-state NMR spectra focusing on the Gly15-Gln16-Gly17 region. (a) PISEMA spectrum of selectively "unlabeled" Pf1 with resonances from only the Asp, Glu and Gln residues. (b)  $^{15}$ N NMR spectrum of an unoriented sample of [ $^{15}$ N]Gly-labeled Pf1.

applied to fd coat protein in lipid bilayers<sup>22</sup> and virus particles.<sup>20</sup> This approach takes advantage of the scattering of each type of residue throughout the sequence and the symmetry inherent in helical structures. Nearly all (41 out of 46) of the residues in the protein are accounted for in the spectra of the nine selectively <sup>15</sup>N-labeled samples shown in Figure 1(d) and the spectrum of the selectively "unlabeled" sample shown in Figure 2(a). The polarity index slant angle (PISA) wheel<sup>23,24</sup> (Figure 1(b) and (c)) and dipolar wave<sup>25,26</sup> (Figure 3) patterns evident in the experimental data were analyzed iteratively, simultaneously assigning the resonances and determining an initial model of the structure.<sup>22,23,27</sup> Using the unique Arg44 and assigned Met19 resonances<sup>28</sup> as anchor points, the sequential assignment of most of the amide resonances from long stretches of helix (residues 7-14 and 18-46) could be obtained by inspection of the wheel-like patterns shown in Figure 1. Some assignments were confirmed by <sup>2</sup>H<sub>2</sub>O exchange experiments (Figure 1(b)).

Although it was well established that most of the coat protein is helical in the bacteriophage particles, the structure of the N-terminal portion was less certain, having been variously described as either helical (PDB entry 1IFN)5,8,29 or unstructured (PDB entry 2IFN)5,30 from poorly defined intensity in fiber diffraction electron density maps.<sup>30</sup> The marked deviation of the structure of the N-terminal region from a continuous helix is illustrated in Figure 1(c). The resonances from the helical regions of the protein appear as PISA wheels, and starting with that from Arg44 near the C terminus follow in a regular counterclockwise fashion with an arc of 100° between adjacent residues (except near Q16) down to Ala7, as expected for an  $\alpha$ -helix. At residue 6 the sense of rotation changes, and the resonances from the N-terminal residues are spread over a broad range of frequencies in a pattern inconsistent with their being in an  $\alpha$ -helix. These resonances correspond to a unique structural feature at the N terminus of Pf1 coat protein.

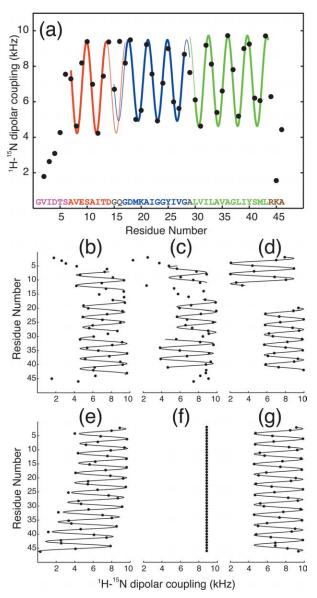


Figure 3. Dipolar waves derived from the spectra in Figure 1. (a) Experimental dipolar waves for the three helical segments of the protein; Ala7-Asp14 (red), Gly17-Gly28 (blue), and Leu30-Leu43 (green). (b) Same as (a), for ease of comparison. (c) Simulated dipolar wave from Pf1 coat protein X-ray fiber diffraction structure 1QL1. (d) Approximate dipolar wave simulated from the description of the Pf1 coat protein neutron diffraction structural model. (e) Simulated dipolar wave for a continuous, gently curved α-helix, oriented at a small angle from vertical, as expected in the intact phage particle. (f) Simulated dipolar wave for a straight ideal  $\alpha$ -helix oriented with the long axis parallel with the magnetic field. (g) Simulated dipolar wave for a straight ideal α-helix with the long axis oriented at an angle of 21° away from the magnetic field.

The remaining residues, Gly15, Gln16, and Gly17, are in a region of the protein that has been at the center of conflicting structural interpretations of the X-ray fiber and neutron diffraction data. The X-ray fiber diffraction data for this region of the protein were interpreted as a continuous,

slightly distorted helix (PDF entry 1QL1).<sup>5</sup> In contrast, distance measurements from neutron diffraction data were interpreted as being inconsistent with a continuous helix between residues 12 and 18, and led to the proposal of a non-helical surface loop to reconcile the data with the adjacent residues being in helices.8 The NMR data also show evidence of structural complexity in this region of the protein. The solid-state NMR spectrum of an unoriented sample of selectively [15N]Gly-labeled Pf1 previously described, 12 and reproduced as part of this study (Figure 2(b)), has a significant amount of narrow resonance intensity at the isotropic frequency (120 ppm) superimposed on what is otherwise a rigid-lattice powder pattern. (The increased intensity at the downfield edge of the powder pattern is a distortion due to the unavoidable partial alignment of the virus particles in a high magnetic field.) This spectrum indicates that most of the glycine residues are immobile and one or more of them are mobile and unstructured on the 10<sup>-4</sup> second timescale of the [<sup>15</sup>N]amide chemical shift interaction. There are seven glycine residues in the protein, and the spectrum from an aligned sample of [15N]Gly-labeled Pf1 shown in Figure 1(d) contains seven resonances; significantly, the same spectrum also contains isotropic intensity from mobile glycine residues at a lower contour level. To identify the backbone amide resonance from Q16, an "unlabeling" experiment was performed. A sample of Pf1 bacteriophage was prepared using [15N]ammonium sulfate based minimal media to which all of the [14N]amino acids, except Q, D, and E had been added. The solution NMR spectrum of the coat protein solubilized in micelles contains only the expected five amide resonances, with a low level of background signals from isotopic scrambling. Therefore, the PISEMA spectrum of the aligned sample (Figure 2(a)) has resonances for all Q, D, and E residues, and an intense signal probably arising from a side-chain. Taken together these data indicate that some of the coat protein subunits have a unique structure and some have mobile, unstructured residues in this region. Molecular dynamics simulations<sup>16,18</sup> suggest that this segment has unusually high flexibility. Residues in this region are mobile in the membrane-bound form of the protein,12,14,17 and they function as the "hinge" for the large change in tertiary structure that accompanies viral assembly. 12,19

## Structure of the Pf1 major coat protein in bacteriophage particles

Dipolar waves describe the periodic variation in the magnitudes of dipolar couplings in the backbone of a protein as a function of residue number, for residues in an  $\alpha$ -helix these are sine waves with a period near 3.6. Further, the parameters obtained when the dipolar couplings are fit to a sinusoid, in particular the amplitudes and phases, can be used to calculate the tilt angle of helical segments with

respect to the direction of the applied magnetic field (and in this case the long axis of the filament) in completely aligned samples.<sup>26,31</sup> The fits of sine waves to the experimental data in Figure 3(a) show that Pf1 coat protein in bacteriophage particles has three distinct helical segments: Ala7-Thr13 (red), Met19-Gly28 (blue), and Leu30-Leu43 (green). Although the amplitudes indicate that the tilt angles of the helical segments are similar (20°-22°), it is apparent that there are changes in phases near Gln16, Ala29, and at the termini of the protein. This is illustrated in Figure 3(a) by continuing the sine waves (thin lines) beyond the regions where they fit well to the experimental data (thick lines). The waves fit to the experimental data (Figure 3(b)) are substantially different from those simulated for the X-ray<sup>5</sup> fiber (Figure 5(c)) and neutron<sup>8</sup> (Figure 3(d)) diffraction structural models. The Dipolar waves for a continuous, gently curving  $\alpha$ -helix<sup>32</sup> (Figure 3(e)), a straight ideal  $\alpha$ -helix parallel with the magnetic field (Figure 3(f)), and a straight ideal helix tilted 21° away from the magnetic field (Figure 3(g)) are not consistent with any of the structural models based on experiments.

The backbone structure of Pf1 coat protein was calculated from the experimental data in Figure 1 and Table 1 by structural fitting.21 The Ramachandran angles for each residue were restrained independently to identify solutions as near as possible to ideal helical values of  $\Phi = -65^{\circ}$ ,  $\Psi = -40^{\circ}$ , but given an allowed range of  $\pm 50^{\circ}$ . Uncertainty in the position of resonance peaks was estimated to be 1 ppm in the chemical shift dimension and 76 Hz in the dipolar dimension. A larger source of potential error is the residue-toresidue variations in the [15N]amide chemical shift tensors. This was taken into account by allowing the magnitudes of the principal values of the chemical shift tensors to vary by  $\pm 7$  ppm between residues and individual structure calculations. Taking into account this range of experimental errors and uncertainties in the chemical shift tensors, the mean RMSD value to the average for 116 calculated structures was less than 0.6 A. Direct calculation of structure and structural fitting based on a set of orientationally dependent frequencies measured in PISEMA spectra yield a unique protein structure. Therefore, in order to obtain an estimate of the accuracy, we perform multiple structural fits with the fitting parameters, e.g. the principal values of the chemical shift tensors, varied within ranges selected to sample all plausible solutions consistent with the experimental data available. The mean RMSD value to the average structure of the ensemble is then calculated. The RMSD value of <1 Å is indicative of a well-determined backbone structure. This is supported by other indicators. For example, Dipolar waves serve as reliable qualitative indicators of structure, and only structures determined by X-ray crystallography or solution NMR spectroscopy with RMSD values <2 A have helices that

**Table 1.** Values used to calculate the solid-state NMR structure of Pf1, 1PJF

Resid	ue	<sup>15</sup> N (ppm)	<sup>1</sup> H- <sup>15</sup> N dipolar coupling (kHz)
G	1		
V	2	164	1.79
Ī	3	91	2.64
D	4	90	3.09
T	5	102	4.26
S	6	164	7.56
A	7	171	7.29
V	8	203	4.63
E	9	221	8.18
S	10	196	9.38
A	11	163	6.98
I	12	188	4.22
Т	13	218	7.44
D	14	183	9.38
G	15	154	6.70
Q	16	207	9.40
G	17	221	8.18
D	18	224	9.49
M	19	175	5.01
K	20	211	5.52
A	21	210	9.22
I	22	164	7.56
G	23	180	4.92
G	24	213	7.05
Y	25	191	8.99
I	26	167	5.99
V	27	191	5.63
G	28	204	8.67
A	29	182	7.64
L	30	177	6.11
V	31	205	4.62
I	32	219	9.18
L	33	174	8.12
A	34	192	5.40
V	35	215	6.60
A	36	209	9.73
G	37	159	7.81
L	38	213	5.19
I	39	220	9.00
Y	40	189	9.25
S	41	183	6.21
M	42	221	6.06
L	43	204	9.70
R	44	143	6.29
K	45	143	1.56
A	46	201	4.43

are well fit by sine waves.<sup>26</sup> The Dipolar waves in Figure 3 are extremely well fit by sine waves with periods of 3.6 residues, and are consistent with the RMSD calculation in showing that the backbone structure of Pf1 coat protein is well determined by this method.

One of the most distinctive features of the structure determined by solid-state NMR spectroscopy is that the N-terminal segment of the protein forms a six-residue "double hook". Previous descriptions of these N-terminal residues in the structures based on X-ray fiber diffraction have included helical (PDB entry 1IFN)<sup>29</sup> and extended, unstructured (PDB entries 2IFN and 1QL1)<sup>5,30</sup> conformations. The individual structure with an RMSD value nearest to the average is shown in stereo view in Figure 4(a), and has been deposited as model 1 of PDB entry 1PJF. By itself, this structural family, which shows a slightly distorted helix

with an inward-directed bulge in the Gly15-Gln16-Gly17 region, does not account for the isotropic intensity observed in the spectra of the [¹⁵N]Gly-labeled samples. Assuming that extensive interactions within the whole virion lattice fix the relative positions and orientations of Asp14 and Asp18, and this supported by the lack of doubling or other distortions in the resonances, a search was made over all Ramachandran space for additional structural solutions that would link Asp14 and Asp18. The original right-handed distorted helical solution was found, and a corresponding left-handed, outward directed "bulge" structural family was also identified, one member of which is shown in stereo view in Figure 4(b).

## Structural model of the Pf1 bacteriophage particle

The resonance frequencies are very sensitive to the two Euler angles between  $^1H^{-15}N$  bonds and the magnetic field that describe the tilt and the twist of a helix, but are insensitive to the third Euler angle (rotation around the magnetic field axis), inversions, and translations. Consequently, information about the relative positions of the subunits in the virus particle is limited to confirming that each protein subunit has the same unique conformation (except for the short mobile segment) and is placed symmetrically such that the line widths and other spectral properties are indistinguishable from those of a well-ordered single crystal.

The resonance assignments shown in Figure 1 are used in the calculation of the atomic-resolution protein structure. They are consistent with the data from selectively  $^{15}\text{N-labeled}$  samples as analyzed with the shotgun approach based on the periodicity of an  $\alpha\text{-helix}$  observed in PISA wheels and Dipolar waves. However, it is also possible to implement structural fitting  $^{21}$  in an "assignment-free" manner, and this yielded a very similar structure with an RMSD value to the average of all the resulting structures of 1.6 Å.

symmetry and distance relationships derived from X-ray fiber and neutron diffraction studies were used to construct a model of the virus particle. The symmetry was 27 copies of the monomer in five turns (in a right-handed sense) around the virion axis, with a rise of 2.9 A per monomer.5 The internal rotation and radial placement of monomers was achieved by a combination of fitting to the neutron diffraction data and energy minimization. In the neutron diffraction studies, the average position of multiple (two to seven) deuterium labels incorporated at each of 15 different amino acid positions was measured.7,8 To incorporate these distance constraints, side-chains were added to the backbone of the initial model of the coat protein using the program SCWRL,33 and the average positions of the deuterium pseudoatoms in each of the deuterated amino acid sidechains were estimated. The positions of the pseudo-atoms were minimized with respect to the

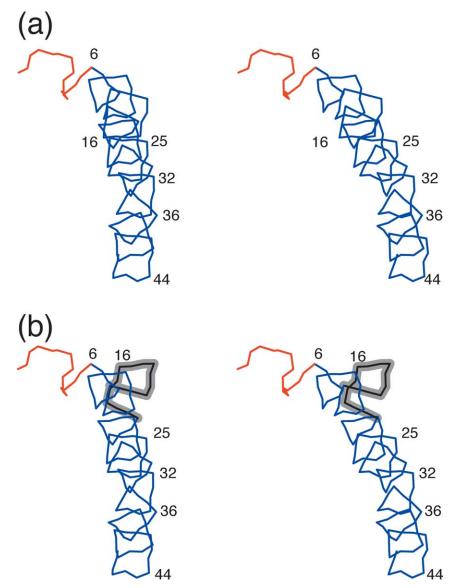
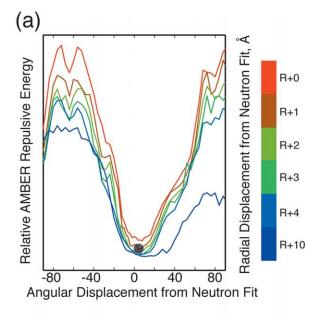


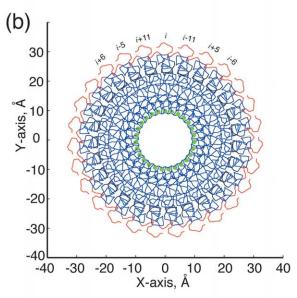
Figure 4. Protein structures calculated from spectra in Figure 1 (data in Table 1) shown in stereo view. (a) Structure nearest to the average from a family of 116, which together have a mean RMSD value to the average structure of  $<0.6\,\text{Å}$ . (b) Alternative "bulge" structure to account for the observation of isotropic resonance intensity in spectra of both unoriented and aligned samples of [ $^{15}$ N]Gly-labeled Pf1. Except for the region of enhanced mobility between Asp14 and Asp18, shown with the black and gray lines, the rest of the structure is the same as in (a).

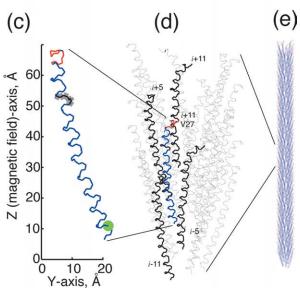
neutron diffraction data by translation in the *XY* plane and rotation around the NMR-insensitive Euler angle, and then the fiber diffraction symmetry was applied to obtain the 27 subunit capsid-repeating unit.

The resulting model of a Pf1 bacteriophage particle, which has an outer diameter of approximately 60 Å and a narrow open bore for the single-stranded DNA in the middle, is generally similar to the fiber diffraction model presented in Figure 12(c) of Welsh *et al.*<sup>34</sup> A large number of 27 monomer capsid structures in which the internal rotation of each monomer was varied over the range  $-92^{\circ}$  to  $+92^{\circ}$  (relative to the rotation that best fit the neutron diffraction data) and the radius incremented in steps of 1 Å were generated. The relative repulsive AMBER energies of the various

configurations were then calculated using the program SCWRL.33 As indicated in Figure 5(a), a small rotation of +4° and radial expansion of 1 Å slightly improved the energy of the model, and is incorporated in PDB entry 1PJF. To eliminate end effects the corresponding 81 monomer assembly was processed with SCWRL, then 27 monomers on each end were removed, leaving the 27 monomer symmetry repeat. Hydrogen atoms were added using the program MOLMOL.35 A backbone-only view down the long axis of the virion model is shown in Figure 5(b), and side views in the context of increasingly large portions of the virion coat are shown in Figure 5(c)-(e). Each monomer at a given position i is closely packed with neighbors at positions  $i \pm 5$ ,  $i \pm 6$ , and  $i \pm 11$ , resulting in extensive hydrophobic interactions







among protein subunits. The N-terminal double hook of monomer i approaches monomer i + 11near residues 23-29, which are a relatively unencumbered GGXXXG sequence. In monomer i, the Asp14-Asp18 region is nearest to the N terminus of monomer i - 6, and there are extensive contacts along opposite faces of i by monomers i-11 and i + 5. In the C-terminal region, there are extensive contacts along opposite faces of i by monomers i-11 and i-5. The coincidence of the N-terminal hook of monomer i and the kink at Ala29 in monomer i + 11 provides structural justification for the presence of a kink in the predominantly helical structure of the protein. The side-chain of Arg44 is directed inward, as expected for interactions with DNA, as well as the side-chain for Tyr40. Although additional information about the tyrosine side-chain orientations from polarized Raman and FTIR spectroscopy has recently become available, 10 this was not explicitly incorporated into the model.

#### **Discussion**

The structure of the Pf1 coat protein monomer (Figure 4(a) and (b)) has several distinctive features. These include the N-terminal double hook (Gly1-Ser6), a region with enhanced mobility including Gly15-Gln16-Gly17, a kink near Ala29, and a slightly unraveled C terminus (Arg44-Lys45-Ala46). These features delimit three nearly ideal straight helical segments (Ala7-Thr13, Met19-Gly28, Leu30-Leu43), each of which has a slightly different tilt angle approximately 20° away from the filament axis. A source of controversy in this field has been differing structural interpretations for the Gly15-Gln16-Gly17 region based

Figure 5. Pf1 major coat protein in the context of a bacteriophage particle. (a) Relative AMBER repulsive energy of 27 monomer assemblies, in which the monomers have been rotated and translated relative to the best fit to the neutron diffraction data. The presence of an energy minimum supports the neutron diffraction data, and identifies radial and angular parameters for constructing the virion (shown with the gray dot). (b) The 27 monomer phage backbone model, viewed down the phage long axis. The X-ray fiber diffraction symmetry of 27 monomers in five turns, with a rise of 2.9 Å/monomer, has been applied. The closest neighbors which interact with a given monomer i ( $i \pm 5$ ,  $\pm 6$ , and  $\pm$  11) are indicated. In this panel and in (c)–(e), the N terminus Gly1 ( $C^{\alpha}$ )-Ser6 ( $C^{\alpha}$ ) is shown as red, the Asp14  $(C^{\alpha})$ -Gly17  $(C^{\alpha})$  potentially mobile region is shown as black with a gray outline, and the green dot indicates the position of Arg44  $C^{\alpha}$ . (c) Pf1 coat protein monomer backbone, side view. (d) Relative positions of neighboring monomers (in heavy black), in a side view of one symmetry repeat unit. The N-terminal double hook of monomer i, near the kink at Ala29 in monomer i + 11 is shown. (e) The 270 monomer phage backbone model, side view. This represents less than 4% of the length of the entire bacteriophage particle.

on X-ray fiber or neutron diffraction data. The solid-state NMR data show that this region is uniquely structured in one population of coat proteins, and mobile in another. The structure determined by NMR is in agreement with the principal finding from the neutron diffraction studies,8 which is that there is a break between helical segments centered near residue 16. The X-ray fiber diffraction data have been interpreted<sup>4</sup> in terms of this protein consisting of "...a single, gently curved stretch of  $\alpha$ -helix..." in the bacteriophage particles. Although apparently in conflict with the results obtained in this solid-state NMR study, the most highly refined X-ray fiber diffraction structure (PDB entry 1QL1) is actually in quite good agreement with the structure determined by NMR (PDB entry 1PJF) with a backbone RMSD value of 1.9 Å for residues 7-46, and an even smaller backbone RMSD value of 1.1 Å for the C-terminal residues 18-46. The notable deviations are in the N-terminal residues where both the X-ray fiber and neutron diffraction data have been interpreted in terms of quite different conformations. It is conceivable that the structural differences previously observed using different methods resulted from differences in the samples,36 especially the hydration of fibers versus solution.

The structure of the N-terminal region has been described as a continuation of the helix from Ala7 and as a random unstructured extension due to a "poorly defined"<sup>30</sup> electron density map for this region; in contrast, the solid state NMR data show that residues 2–6 form a completely immobile and uniquely structured double hook that wraps around the outer perimeter of the phage, just above the Ala29 kink of a neighboring subunit. There are several potential biological roles for this structural feature. As it wraps around the surface of the phage, it may contribute to holding neighboring coat protein monomers together, and thus contribute to the stability of the particle. It may have a role in one of the key events of the viral lifecycle, including infection or extrusion of new particles. Mutagenesis studies directed at the N-terminal region may give insights into the roles of these residues in phage viability, stability, and infectivity. The structure of this N-terminal region of the coat protein is of particular interest in the context of phage display where peptide sequences are inserted near the N terminus of the major coat protein and exposed on the surface of the bacteriophage particles. While the fd and M13 phage display systems are better characterized, there may be some advantages to a Pf1 phage display system, since the structured N terminus of Pf1 coat protein provides a more rigid scaffold for the displayed peptides.

A recent hypothesis<sup>5</sup> in fiber diffraction studies of Pf1 is that the coat protein subunits are not all arranged symmetrically in the bacteriophage particles, but rather possess slightly different orientations within subgroups of three monomers. Solid-state NMR of aligned samples is well suited

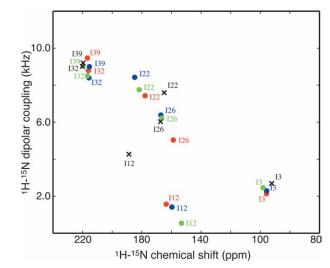


Figure 6. Simulated spectra of a selectively [15N]isoleucine-labeled sample from the three differently oriented coat protein monomers in the fiber diffraction structural model (PDB entry 1QL2), comparing the "groupings of three" symmetry model. Simulated peaks are shown as dots of 300 Hz diameter in red, blue, or green (for the three different coat protein monomers), and the positions corresponding to the solid-state NMR experimental data are shown as black X symbols. The spread in the peak positions of the colored dots for a given amino acid is not consistent with the individual peaks seen in the experimental solid-state NMR data.

to address this issue. As seen in the spectra in Figure 1, each amide site gives rise to a single, well-defined resonance, with line widths for individual peaks of around 250 Hz in the chemical shift dimension, and 500 Hz in the dipolar dimension. This indicates an exquisitely high level of symmetry; each of the 7000 copies of the coat protein in a bacteriophage particle has the same structure and orientation with respect to the magnetic field. In contrast, the spectra in Figure 6 simulated for each of the monomers in the fiber diffraction structural model with groups of three differently oriented coat proteins (PDB 1QL2) show multiple resonance peaks for a given backbone amide site which range in frequency significantly beyond the observed line widths, and this is clearly inconsistent with the experimental NMR data.

Dramatic structural changes occur when the coat protein undergoes the transition from the bacterial membrane to the phage particle during assembly. Comparison of the structure of the coat protein in bacteriophage particles to the membrane-bound form characterized by solution NMR of micelle<sup>17</sup> and solid-state NMR spectra of bilayer<sup>12</sup> samples reveals major differences in both structure and dynamics. The membrane-bound form of the coat protein has a C-terminal *trans*-membrane helix (Gly24-Arg44) oriented nearly perpendicular to an in-plane amphipathic helix (Ser6-Thr13), while in the intact bacteriophage particle, these same helices are aligned approximately parallel with

each other. In micelles, the membrane-bound form of the protein has a mobile region between residues Asp14 and Gly23. 12,17 A remnant of this mobile "hinge" region is retained in the protein in bacteriophage particles; however, it is much smaller, comprising only residues Gly15-Gln16-Gly17. In the membrane-bound form of the coat protein, the N and C-terminal regions are mobile; in contrast, the same terminal residues are structured and immobile in the bacteriophage particle. This is supported by our earlier <sup>2</sup>H NMR studies, <sup>12</sup> which showed that Thr5 and Leu43 are mobile in bilayer samples, but immobile and structured in the bacteriophage.

fd and Pf1 are representatives of class I and class II bacteriophages, respectively. While the genomes of these two bacteriophages are approximately the same size, fd is less than half the length of Pf1, which reflects their nucleotide to coat protein ratios  $(2.4:1 \ \textit{versus} \ 1:1)$ . The coat proteins differ only slightly in length (50 versus 46 amino acid residues), but share little primary sequence similarity and are arranged in the bacteriophage particles with different symmetries. Both coat proteins are comprised primarily of  $\alpha$ -helical segments tilted approximately 20° away from the filament axis; however, there are several notable differences in their structures. Unlike Pf1, the N-terminal region of fd is mobile and unstructured at temperatures as low as 0 °C. There is no evidence of an internal mobile region in fd corresponding to the Gly15-Gln16-Gly17 region of Pf1. The residues in the hinge region of fd coat protein appear to integrate seamlessly into a single  $\alpha$ -helix, in contrast, those in Pf1 coat protein provide a break between two helical segments. While fd and Pf1 both have a distinct deviation from an ideal helix toward the C terminus (near Ala29 in Pf1 and Ile39 in fd), the position and extent of the kink is different. In Pf1 coat protein, there is a fairly subtle change in tilt angle, identified as a change in the phase of the dipolar wave (Figure 3(a)), without a significant change in the amplitudes of the waves. In fd coat protein, the kink is more pronounced (ca 5°) and results in a significant change in amplitude of the waves.20 Thus, although Pf1 and fd are similar nucleoprotein assemblies, they have significant structural differences at both the microscopic and atomic levels of resolution.

#### Methods

#### Pf1 sample preparation

For uniformly  $^{15}\text{N}$ -labeled bacteriophage, the minimal growth media was comprised of the following: 11 g/l Na<sub>2</sub>HPO<sub>4</sub>, 3 g/l KH<sub>2</sub>PO<sub>4</sub>, 0.5 g/l NaCl, 1 g/l MgSO<sub>4</sub>–7H<sub>2</sub>O, 0.025 g/l CaCl<sub>2</sub>–2H<sub>2</sub>O, 10 g/l dextrose, 0.1 g/l thiamine hydrochloride, 10 ml/l 100X MEM vitamin solution (Sigma, M6895), and 1 g/l [ $^{15}\text{N}$ ]ammonium sulfate. For selectively labeled samples, one  $^{15}\text{N}$ -labeled amino acid (Cambridge Isotope Laboratories) was added at a concentration of 0.1 g/l, and the other 19 unlabeled

amino acids were added at higher (>0.2 g/l) concentrations. The final pH of the media was adjusted to approximately pH 7.4. Pseudomonas aeruginosa (ATCC, 25102) was grown in 21 baffled flasks to  $A_{600}$  of 0.337<sup>37</sup>, and then infected with Pf1 phage (ATCC, 25102-B1) at a high multiplicity of infection. Growth was allowed to continue from four hours to overnight, after which the cultures were centrifuged at low speed, the supernatant retained, and the phage precipitated by addition of polyethylene glycol (PEG)-8000 (40 g/l) and NaCl (29.2 g/l) with cooling overnight at 4 °C. The precipitated phage were collected by centrifugation, resuspended in TE buffer<sup>37</sup> (pH 8), centrifuged to remove any remaining bacteria, re-precipitated with 4% PEG-8000, 0.5 M NaCl, and re-suspended overnight in 5 mM sodium borate buffer (pH 8). The concentration was adjusted to approximately 1 mg/ml phage and then mixed with solid CsCl (0.4 g per ml of 1 mg/ml phage) and prepared in OptiSeal tubes (Beckman, 362185) for density gradient ultracentrifugation. Centrifugation was performed in a Beckman NVT90 rotor from four hours to overnight at 70,000 rpm and 10 °C. The phage particles form a pale blue band near the center of the tube, and were removed with a short Pasteur pipette and placed into 10 kDa cutoff dialysis tubing. Following dialysis, the phage solution was centrifuged at 50,000 rpm for two hours to obtain a pellet of viscous phage gel (around 90 mg/ml), which was diluted with sodium borate/sodium azide buffer to approximately 50 mg/ml. A final volume of 100-200 μl was placed in 15-16 mm long, 5 mm outer diameter thin wall glass tubes for spectroscopy.

#### NMR spectroscopy

The solid-state NMR experiments were performed on a spectrometer with a wide-bore Magnex 550/89 magnet and a home-built console controlled by a Tecmag Apollo system, and on a spectrometer with a standard-bore Magnex 750/54 magnet and a Bruker Avance console with high-power amplifiers on all channels. Home-built probes with a single horizontal solenoid coil double tuned to the (\frac{15}{N} and \frac{1}{H}) resonance frequencies were utilized. Chemical shift referencing was relative to solid, powdered [\frac{15}{N}]ammonium sulfate as an external reference, and defined as 26.8 ppm. Spectra were acquired with the sample temperature regulated at 30 °C.

The PISEMA and SAMMY pulse sequences were employed as described.  $^{38-40}$  A typical set of acquisition setup parameters was the following. The  $^{1}$ H carrier frequency was set at +2 ppm relative to the water resonance. The  $^{15}$ N carrier frequency was set at 193 ppm, or +166.2 ppm relative to the signal from powdered  $[^{15}$ N]ammonium sulfate. The  $^{1}$ H 90° pulse length was 4.5  $\mu$ s (corresponding to a decoupling field strength of 56 kHz), and the initial spin lock cross-polarization time was 1 ms. During the frequency switched Lee-Goldburg period of the PISEMA, the power on the  $^{15}$ N channel was increased to match the power level on the  $^{1}$ H channel. In this particular example, 128 transients were acquired with 56 points in the indirect  $t_1$  dimension, using a recycle delay of six seconds.

#### Data processing

The experimental data were processed using the program Felix (Accelrys, San Diego, CA). The spectra from selectively <sup>15</sup>N-labeled samples were typically processed with parameters that optimized the signal-to-noise ratio,

at the expense of some resolution; exponential line broadening was applied in both the  $^{15}\mathrm{N}$  chemical shift and  $^{1}\mathrm{H}-^{15}\mathrm{N}$  dipolar coupling dimensions. In spectra of uniformly  $^{15}\mathrm{N}$ -labeled samples, apodization with phase-shifted sine bells in both dimensions was used to optimize resolution. Some processing included a polynomial baseline correction in the  $^{15}\mathrm{N}$  chemical shift dimension and a Lorentz to Gauss transformation in both dimensions.

#### Structure calculations

Fitting of a periodic sine wave to the dipolar coupling data was performed using MATLAB scripts.<sup>25,26,31</sup> Structural fitting was performed using a modified version of a MATLAB script and algorithm as described.<sup>21</sup> Briefly, the fitting algorithm uses both the dipolar coupling and chemical shift data together with restraints on the allowed Ramachandran  $\Phi$  and  $\Psi$  backbone angles and an estimate of the errors in the resonance peak positions to limit the number of structures consistent with the NMR data to a unique family, or small set of related structural families. The peptide plane geometry is assumed to be constant with the values listed below, which provide chemical bond constraints for the fitting. The fitting algorithm was modified to iteratively adjust the allowed  $\Phi$  and  $\Psi$  values independently for each amino acid position to maintain the searched Ramachandran space as near as possible to the ideal helical values  $(\Phi = -65^{\circ}, \ \Psi = -40^{\circ})$ . The tolerance, or peak position error, used in the calculation of PDB structure 1PJF was ±76 Hz for all residues. The bond angles used were the following:  $HN\sigma 33 = 18.5^{\circ}$ ;  $HNC^{\alpha} = 118.2^{\circ}$ ;  $NCoC^{\alpha} = 115.6^{\circ}$ ;  $NC^{\alpha}Co = 110.5^{\circ}$ ;  $C^{\alpha}CoO = 121.1^{\circ}$ ;  $NCoO = 123.2^{\circ}$ . The bond lengths used were the following:  $C^{\alpha}Co=1.52~\text{Å}$ ; CoN=1.33~Å;  $NC^{\alpha}=1.46~\text{Å}$ ; NH=1.07~Å. The  $^{15}N$  chemical shift tensors were the following. For all residues except glycine:  $\sigma 11 =$  $64(\pm 7)$  ppm;  $\sigma 22 = 77(\pm 7)$  ppm;  $\sigma 33 = 222(\pm 7)$  ppm. For glycine residues:  $\sigma 11 = 41(\pm 7)$  ppm;  $\sigma 22 =$  $64(\pm 7)$  ppm;  $\sigma 33 = 215(\pm 7)$  ppm.

To obtain the possible "mobile" solutions for the Gly15-Gln16-Gly17 region, the positions of peptide planes for Asp14 and Asp18 were held fixed and geometrically possible backbone trajectories were calculated starting from the fixed position of the Asp14 peptide plane. The orientations of the residue planes Gly15-Asp18 were propagated by four pairs of Ramachandran angles  $\Phi$  and  $\Psi$ . The RMSD values at the end of the trajectories were minimized with respect to the position of atoms of the Asp18 plane in 1PJF. A solution was accepted if the RMSD value was less than 0.01 Å.

#### Structure coordinates

The coordinates of the solid state NMR structure of the Pf1 major coat protein in bacteriophage particles (structure 1PJF) have been deposited at the RCSB Protein Database and BioMagRes Bank for immediate release upon publication.

#### **Acknowledgements**

We thank M. Mesleh for helpful discussions, and A. Majumdar and G. Chukkapalli at the San Diego

Supercomputer Center for their efforts to parallelize the computer program SCWRL for use on the Blue Horizon supercomputer, as well as R. Dunbrack (Fox Chase Cancer Center) for permission to do so. We also thank L. Makowski and D. Marvin for their interest, assistance, and critiques throughout the thirty-year history of this project, as well as R. Perham for his input. This research made use of computer time at the San Diego Supercomputer Center available through the UCSD Academic Associates program. This research was supported by grants from the National Institutes of Health RO1EB001966, RO1EB002169, and P41EB002031, which supports the Biomedical Technology Resource for NMR Molecular Imaging of Proteins. D.S.T. was supported by a postdoctoral fellowship (F32GM63300) from the National Institutes of Health. The authors declare they have no competing financial interest.

#### References

- Webster, R. (2001). Filamentous phage biology. In Phage Display: A Laboratory Manual (Barbas, C. F., Burton, D. R., Scott, J. K. & Silverman, G. J., eds), pp. 1.1–1.37, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. Curr. Opin. Struct. Biol. 8, 150–158.
- 3. Makowski, L., Caspar, D. L. & Marvin, D. A. (1980). Filamentous bacteriophage Pf1 structure determined at 7 Å resolution by refinement of models for the alpha-helical subunit. *J. Mol. Biol.* **140**, 149–181.
- 4. 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.
- Welsh, L. C., Symmons, M. F. & Marvin, D. A. (2000). The molecular structure and structural transition of the alpha-helical capsid in filamentous bacteriophage Pf1. Acta Crystallog. sect. D, 56, 137–150.
- Liu, D. J. & Day, L. A. (1994). Pf1 virus structure: helical coat protein and DNA with paraxial phosphates. Science, 265, 671–674.
- Stark, W., Glucksman, M. J. & Makowski, L. (1988). Conformation of the coat protein of filamentous bacteriophage Pf1 determined by neutron diffraction from magnetically oriented gels of specifically deuterated virions. J. Mol. Biol. 199, 171–182.
- 8. Nambudripad, R., Stark, W. & Makowski, L. (1991). Neutron diffraction studies of the structure of filamentous bacteriophage Pf1. Demonstration that the coat protein consists of a pair of alpha-helices with an intervening, non-helical surface loop. *J. Mol. Biol.* 220, 359–379.
- Thomas, G., Jr & Murphy, P. (1975). Structure of coat proteins in Pf1 and fd virions by laser Raman spectroscopy. *Science*, 188, 1205–1207.
- 10. Tsuboi, M., Kubo, Y., Ikeda, T., Overman, S. A., Osman, O. & Thomas, G., Jr (2003). Protein DNA residue orientations in the filamentous virus Pf1

- determined by polarized Raman and polarized FTIR spectroscopy. *Biochemistry*, **42**, 940–950.
- 11. Cross, T. A., Tsang, P. & Opella, S. J. (1983). Comparison of protein and deoxyribonucleic acid backbone structures in fd and Pf1 bacteriophages. *Biochemistry*, **22**, 721–726.
- 12. Shon, K. J., Kim, Y., Colnago, L. A. & Opella, S. J. (1991). NMR studies of the structure and dynamics of membrane-bound bacteriophage Pf1 coat protein. *Science*, **252**, 1303–1305.
- 13. Bogusky, M. J., Tsang, P. & Opella, S. J. (1985). Oneand two-dimensional <sup>15</sup>N/<sup>1</sup>H NMR of filamentous phage coat proteins in solution. *Biochem. Biophys. Res. Commun.* **127**, 540–545.
- 14. Schiksnis, R. A., Bogusky, M. J., Tsang, P. & Opella, S. J. (1987). Structure and dynamics of the Pf1 filamentous bacteriophage coat protein in micelles. *Biochemistry*, **26**, 1373–1381.
- Schiksnis, Ř. A., Bogusky, M. J. & Opella, S. J. (1988).
  Secondary structure of filamentous bacteriophage coat protein is preserved in lipid environments. *J. Mol. Biol.* 200, 741–743.
- Tobias, D. J., Gesell, J., Klein, M. L. & Opella, S. J. (1995). A simple protocol for identification of helical and mobile residues in membrane proteins. *J. Mol. Biol.* 253, 391–395.
- 17. Lee, S., Mesleh, M. F. & Opella, S. J. (2003). Structure and dynamics of a membrane protein in micelles from three solution NMR experiments. *J. Biomol. NMR*, **26**, 327–334.
- Tobias, D. J., Klein, M. L. & Opella, S. J. (1993).
  Molecular dynamics simulation of Pf1 coat protein. Biophys. J. 64, 670–675.
- Nambudripad, R., Stark, W., Opella, S. J. & Makowski, L. (1991). Membrane-mediated assembly of filamentous bacteriophage Pf1 coat protein. *Science*, 252, 1305–1308.
- Zeri, A. C., Mesleh, M. F., Nevzorov, A. A. & Opella, S. J. (2003). Structure of the coat protein in fd filamentous bacteriophage particles determined by solid-state NMR spectroscopy. *Proc. Natl Acad. Sci.* USA, 100, 6458–6463.
- 21. Nevzorov, A. A. & Opella, S. J. (2003). Structural fitting of PISEMA spectra of aligned proteins. *J. Magn. Reson.* **160**, 33–39.
- 22. Marassi, F. M. & Opella, S. J. (2003). Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints. *Protein Sci.* **12**, 403–411.
- Marassi, F. M. & Opella, S. J. (2000). A solid-state NMR index of helical membrane protein structure and topology. J. Magn. Reson. 144, 150–155.
- 24. Wang, J., Denny, J., Tian, C., Kim, S., Mo, Y., Kovacs, F. et al. (2000). Imaging membrane protein helical wheels. J. Magn. Reson. 144, 162–167.
- Mesleh, M. F., Veglia, G., DeSilva, T. M., Marassi, F. M. & Opella, S. J. (2002). Dipolar waves as NMR maps of protein structure. J. Am. Chem. Soc. 124, 4206–4207.

- Mesleh, M. F. & Opella, S. J. (2003). Dipolar waves as NMR maps of helices in proteins. J. Magn. Reson. 163, 288–299.
- 27. Marassi, F. M. & Opella, S. J. (2002). Using pisa pies to resolve ambiguities in angular constraints from PISEMA spectra of aligned proteins. *J. Biomol. NMR*, **23**, 239–242.
- 28. Tsang, P. (1984). NMR Studies of the Filamentous Bacteriophage, Pf1. PhD. Thesis, Department of Chemistry, University of Pennsylvania.
- Marvin, D. A. (1990). Model-building studies of Inovirus: genetic variations on a geometric theme. *Int. J. Biol. Macromol.* 12, 125–138.
- 30. Gonzalez, A., Nave, C. & Marvin, D. A. (1995). Pf1 filamentous bacteriophage: refinement of a molecular model by simulated annealing using 3.3 Å resolution X-ray fibre diffraction data. *Acta Crystallog. sect. D*, **51**, 792–804.
- 31. Mesleh, M. F., Lee, S., Veglia, G., Thiriot, D. S., Marassi, F. M. & Opella, S. J. (2003). Dipolar waves map the structure and topology of helices in membrane proteins. *J. Am. Chem. Soc.* **125**, 8928–8935.
- Zhou, N. E., Zhu, B.-Y., Sykes, B. D. & Hodges, R. S. (1992). Relationship between amide proton chemical shifts and hydrogen bonding in amphipathic alphahelical peptides. J. Am. Chem. Soc. 114, 4320–4326.
- Bower, M. J., Cohen, F. E. & Dunbrack, R. L., Jr (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* 267, 1268–1282.
- 34. Welsh, L. C., Symmons, M. F., Sturtevant, J. M., Marvin, D. A. & Perham, R. N. (1998). Structure of the capsid of Pf3 filamentous phage determined from X-ray fibre diffraction data at 3.1 Å resolution. *J. Mol. Biol.* **283**, 155–177.
- 35. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOL-MOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.
- Specthrie, L., Greenberg, J., Glucksman, M. J., Diaz, J. & Makowski, L. (1987). Structural responsiveness of filamentous bacteriophage Pf1: comparison of virion structure in fibers and solution. The effect of temperature and ionic strength. *Biophys. J.* 52, 199–214.
- Sambrook, J. & Russell, D. W. (2001). Molecular Cloning—A Laboratory Manual, vol. A1.7, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 38. Nevzorov, A. A. & Opella, S. J. (2003). A "Magic Sandwich" pulse sequence with reduced offset dependence for high-resolution separated local field spectroscopy. *J. Magn. Reson.* **164**, 182–186.
- Wu, C. H., Ramamoorthy, A. & Opella, S. J. (1994).
  High-resolution heteronuclear dipolar solid-state NMR-spectroscopy. J. Magn. Reson. ser. A, 109, 270–272.
- Ramamoorthy, A., Wu, C. H. & Opella, S. J. (1999).
  Experimental aspects of multidimensional solidstate NMR correlation spectroscopy. *J. Magn. Reson.* 140, 131–140.

Edited by A. G. Palmer III